Learning from 'Binary Labels':

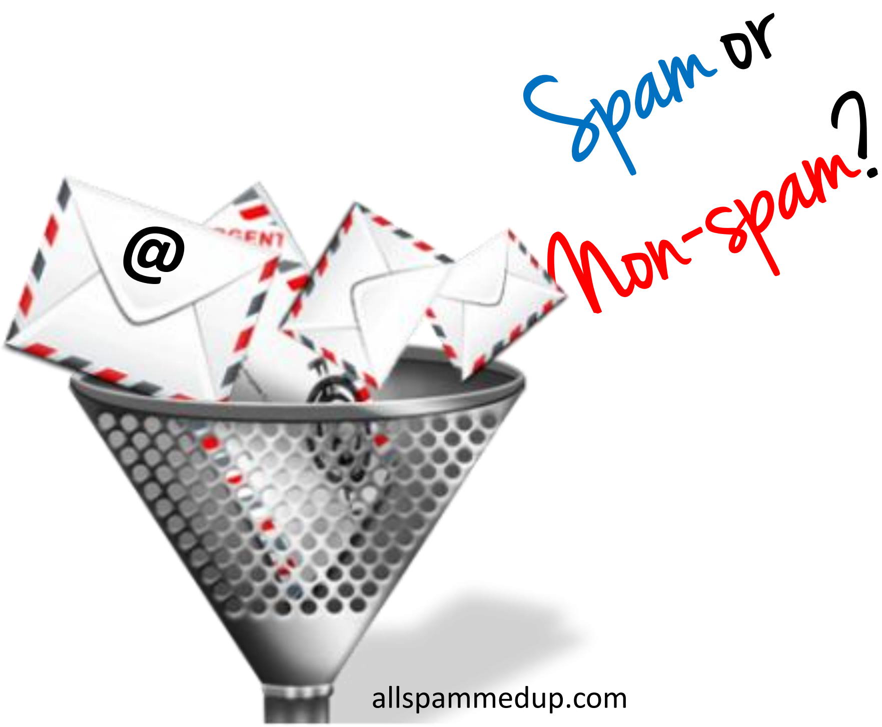
A Plethora of Performance Measures, A Plethora of Algorithms Needed!

Harikrishna Narasimhan

Department of Computer Science and Automation
Indian Institute of Science, Bangalore

Spam or

Non-spam?


Spam or
Non-spam?
@
allspammedup.com


Disease or
No Disease?
pascal-network.org

Spam or Non-spam?

allspammedup.com

Disease or No Disease?

pascal-network.org

fusionsedge.com

Relevant or Irrelevant?

Spam or Non-spam?

allspammedup.com

Disease or No Disease?

pascal-network.org

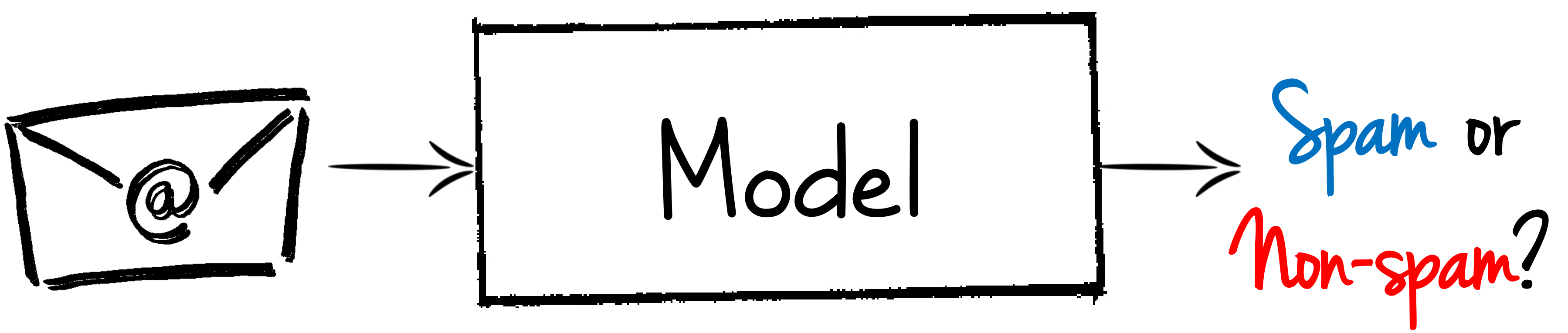
Relevant or Irrelevant?

fusionsedge.com

Click or No Click?

optimum7.com

Standard Binary Classification



Standard Binary Classification

$$\sum_{i=1}^{N} \mathbf{1}\big(y_i \neq h(x_i)\big)$$

Standard Binary Classification

convex upper bound

$$\min_{f \in \mathcal{F}} \left[\sum_{i=1}^{N} \ell\big(y_i, f(x_i)\big) + \lambda R(f) \right]$$

Standard Binary Classification

Logistic Regression

$$\min_{f \in \mathcal{F}} \left[\sum_{i=1}^{N} \ell_{\log}(y_i, f(x_i)) + \lambda R(f) \right]$$

$$\ell_{\log}(y, f) = \log(1 + e^{-yf})$$

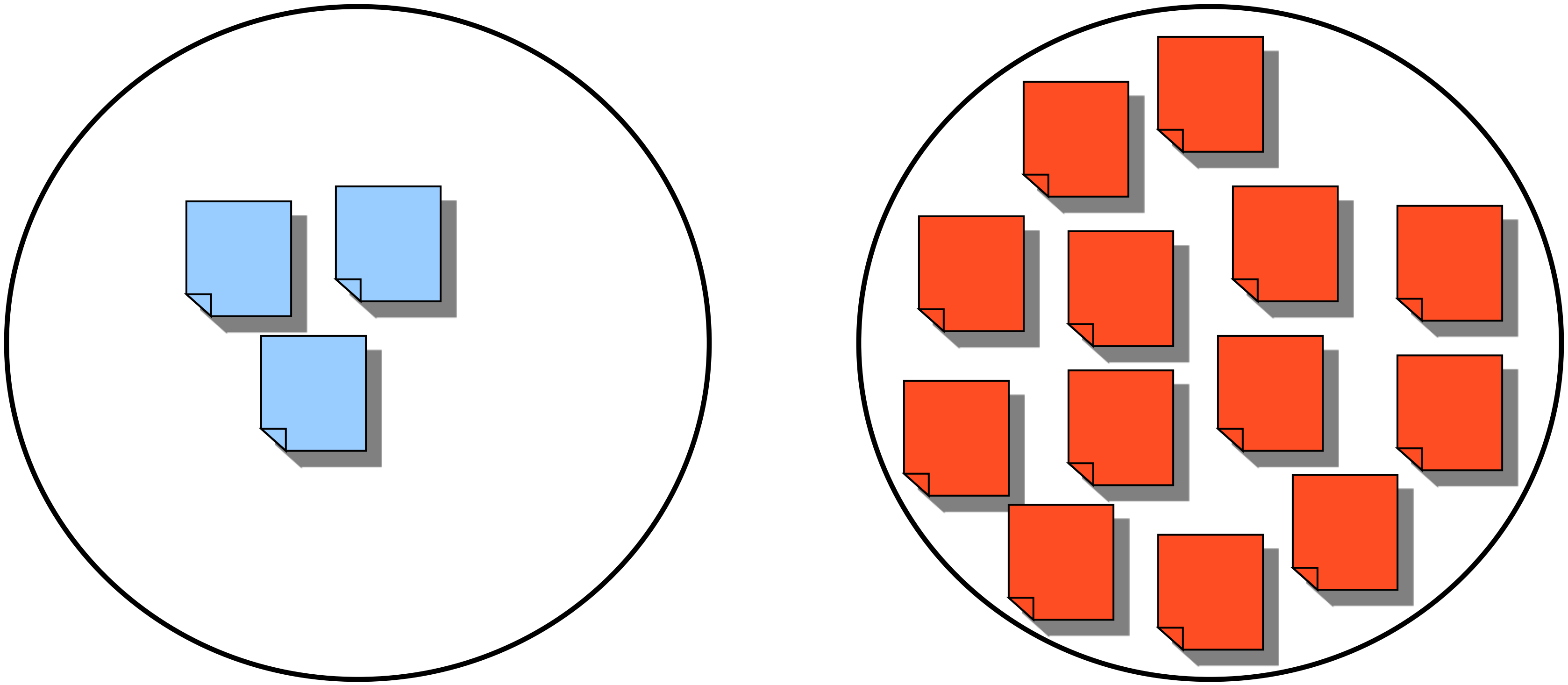
Standard Binary Classification

Support Vector Machine

$$\min_{f \in \mathcal{F}} \left[\sum_{i=1}^{N} \ell_{\text{hinge}}\big(y_i, f(x_i)\big) + \lambda R(f) \right]$$

$$\ell_{\text{hinge}}(y, f) = (1 - yf)_+$$

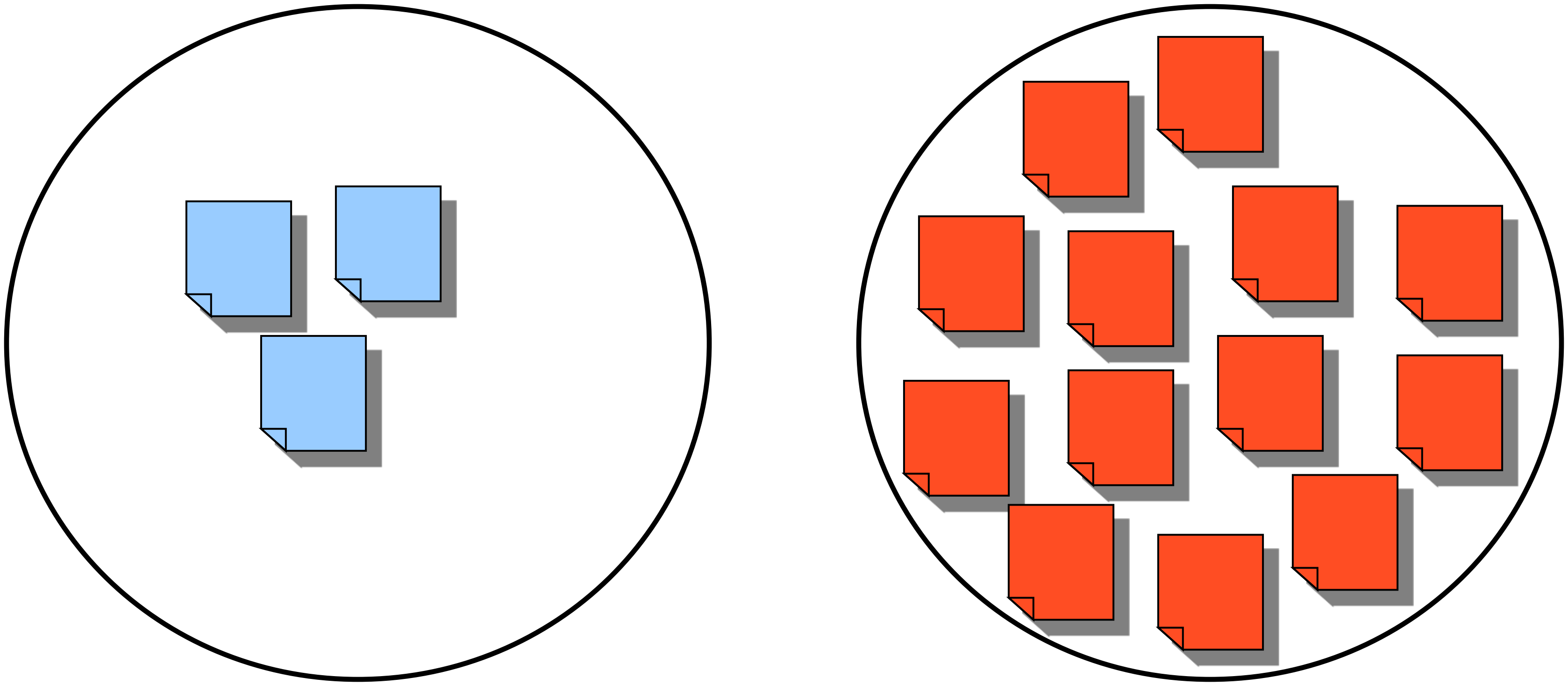
Text Retrieval



Standard classification error ill-suited!

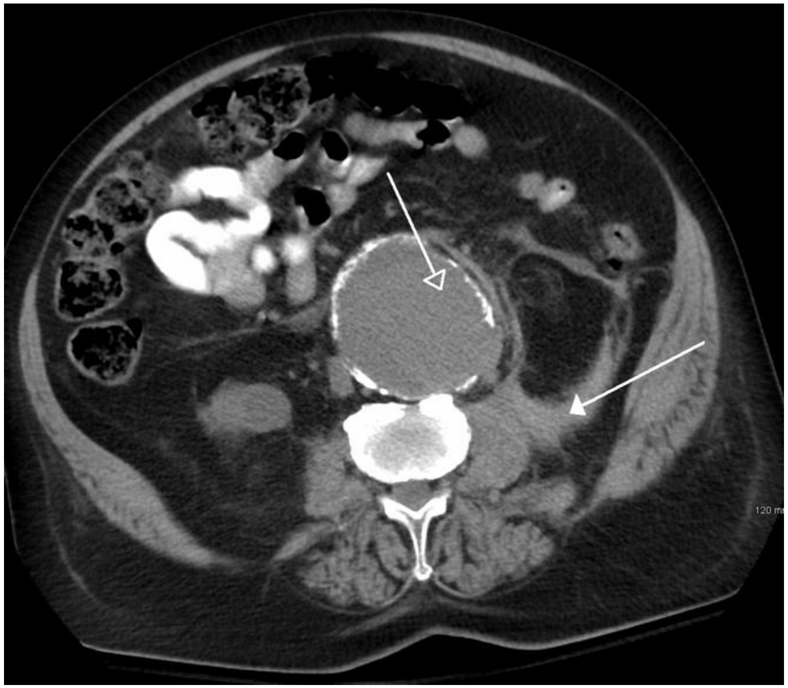
Text Retrieval

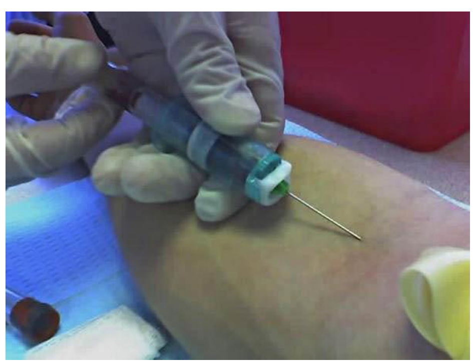
F-measure



Standard classification error ill-suited!

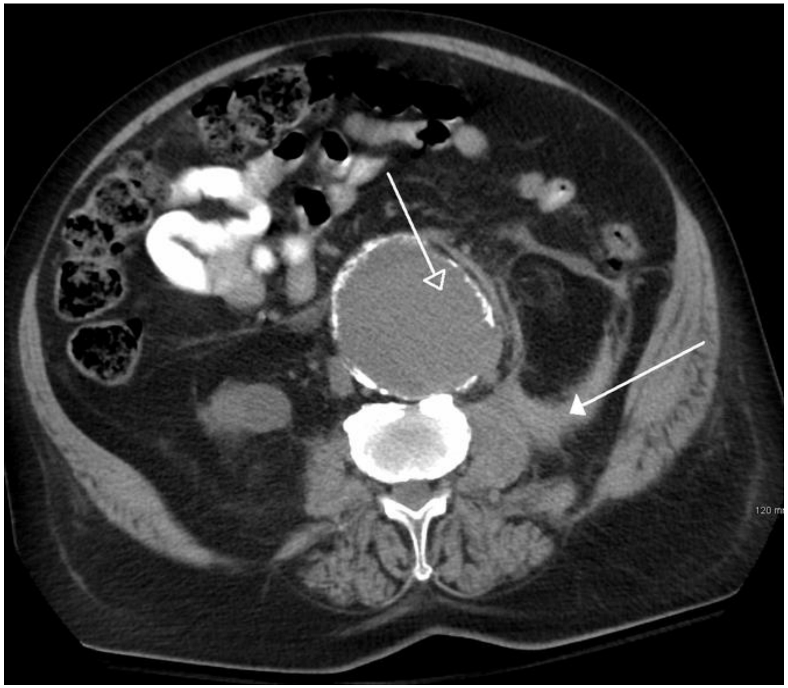
Medical Diagnosis

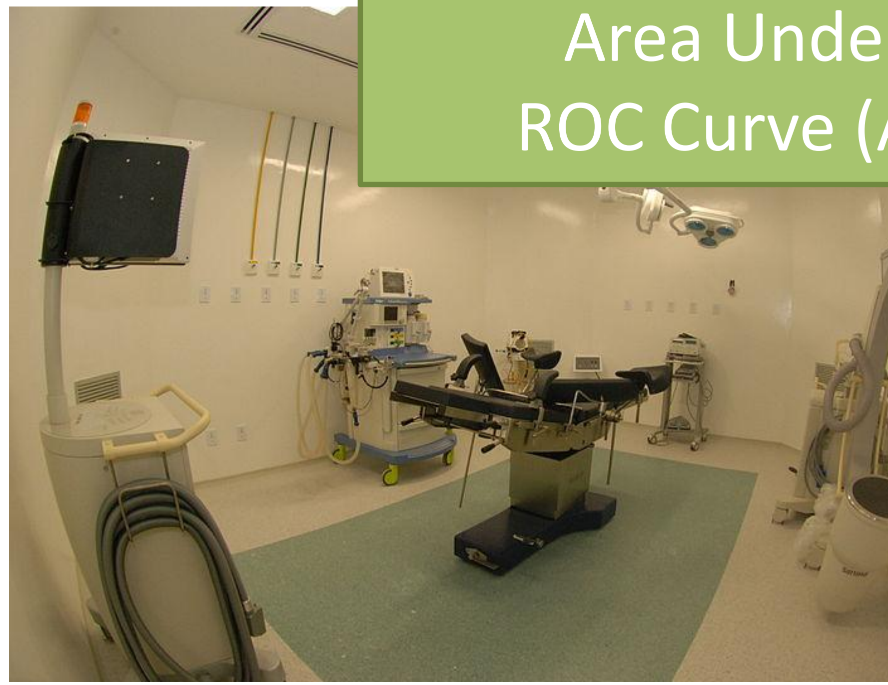




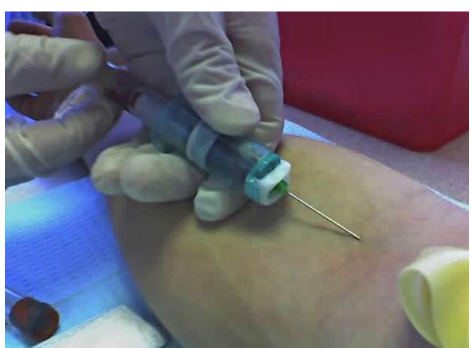
http://en.wikipedia.org/

Medical Diagnosis



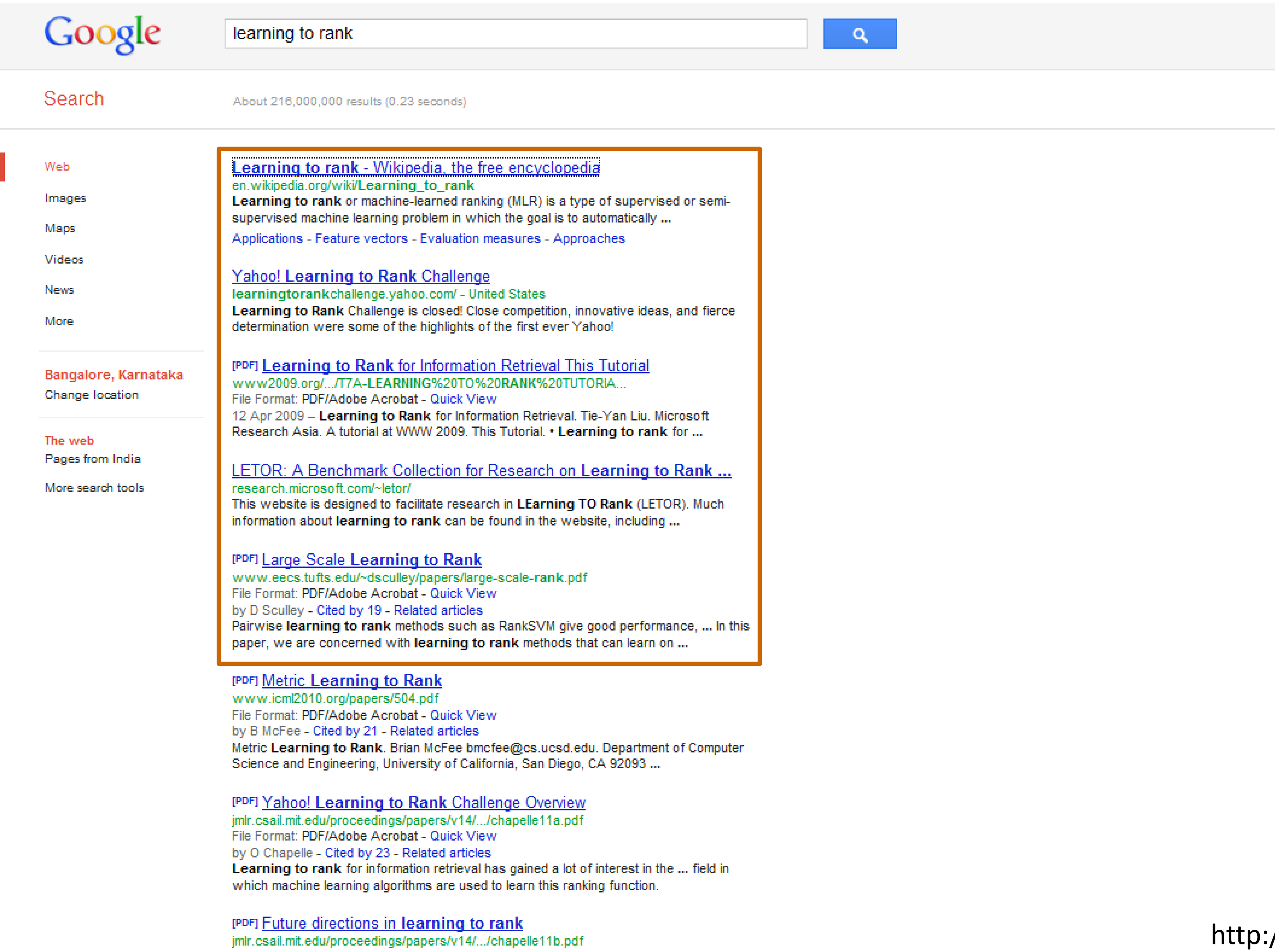


Area Under the ROC Curve (AUC)



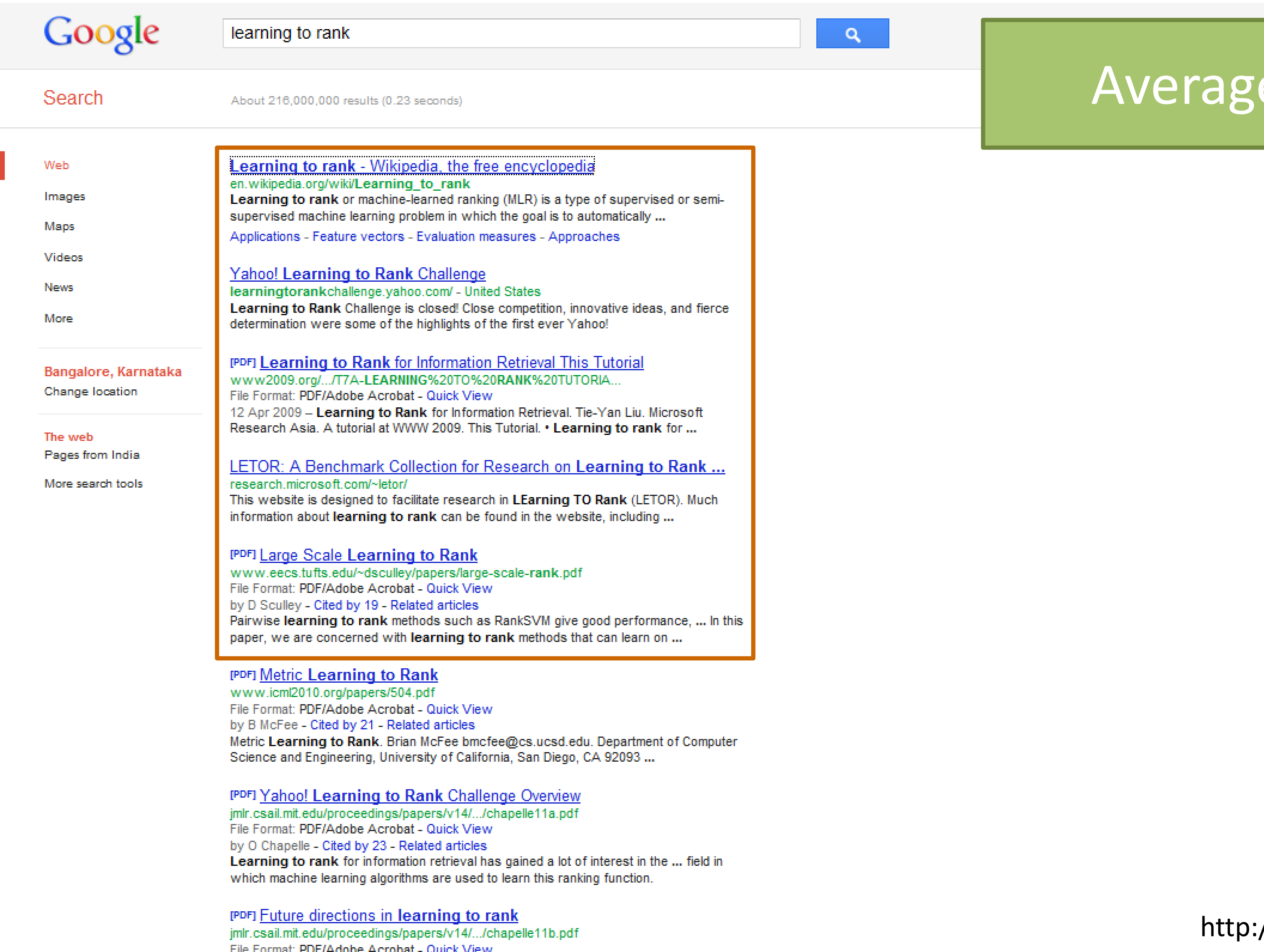
http://en.wikipedia.org/

Information Retrieval



http://www.google.com/

Information Retrieval



Google

learning to rank

Search

About 216,000,000 results (0.23 seconds)

Web
Images
Maps
Videos
News
More

Bangalore, Karnataka
Change location

The web
Pages from India

More search tools

Learning to rank - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Learning_to_rank
Learning to rank or machine-learned ranking (MLR) is a type of supervised or semi-supervised machine learning problem in which the goal is to automatically ...
Applications - Feature vectors - Evaluation measures - Approaches

Yahoo! Learning to Rank Challenge
learningtorankchallenge.yahoo.com/ - United States
Learning to Rank Challenge is closed! Close competition, innovative ideas, and fierce determination were some of the highlights of the first ever Yahoo!

[PDF] Learning to Rank for Information Retrieval This Tutorial
www2009.org/.../T7A-LEARNING%20TO%20RANK%20TUTORIA...
File Format: PDF/Adobe Acrobat - Quick View
12 Apr 2009 – **Learning to Rank** for Information Retrieval. Tie-Yan Liu. Microsoft Research Asia. A tutorial at WWW 2009. This Tutorial. • **Learning to rank** for ...

LETOR: A Benchmark Collection for Research on Learning to Rank ...
research.microsoft.com/~letor/
This website is designed to facilitate research in **LEarning TO Rank** (LETOR). Much information about **learning to rank** can be found in the website, including ...

[PDF] Large Scale Learning to Rank
www.eecs.tufts.edu/~dsculley/papers/large-scale-rank.pdf
File Format: PDF/Adobe Acrobat - Quick View
by D Sculley - Cited by 19 - Related articles
Pairwise **learning to rank** methods such as RankSVM give good performance, ... In this paper, we are concerned with **learning to rank** methods that can learn on ...

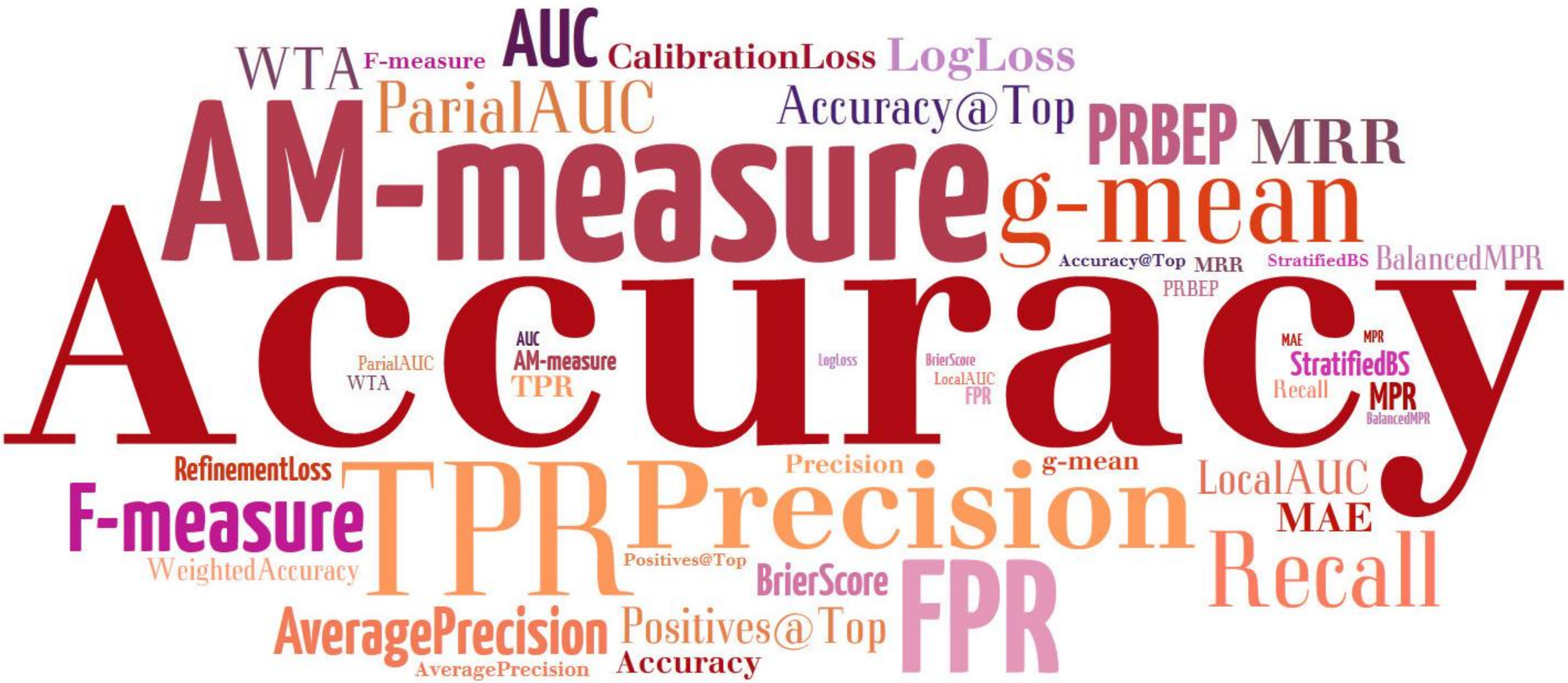
[PDF] Metric Learning to Rank
www.icml2010.org/papers/504.pdf
File Format: PDF/Adobe Acrobat - Quick View
by B McFee - Cited by 21 - Related articles
Metric **Learning to Rank**. Brian McFee bmcfee@cs.ucsd.edu. Department of Computer Science and Engineering, University of California, San Diego, CA 92093 ...

[PDF] Yahoo! Learning to Rank Challenge Overview
jmlr.csail.mit.edu/proceedings/papers/v14/.../chapelle11a.pdf
File Format: PDF/Adobe Acrobat - Quick View
by O Chapelle - Cited by 23 - Related articles
Learning to rank for information retrieval has gained a lot of interest in the ... field in which machine learning algorithms are used to learn this ranking function.

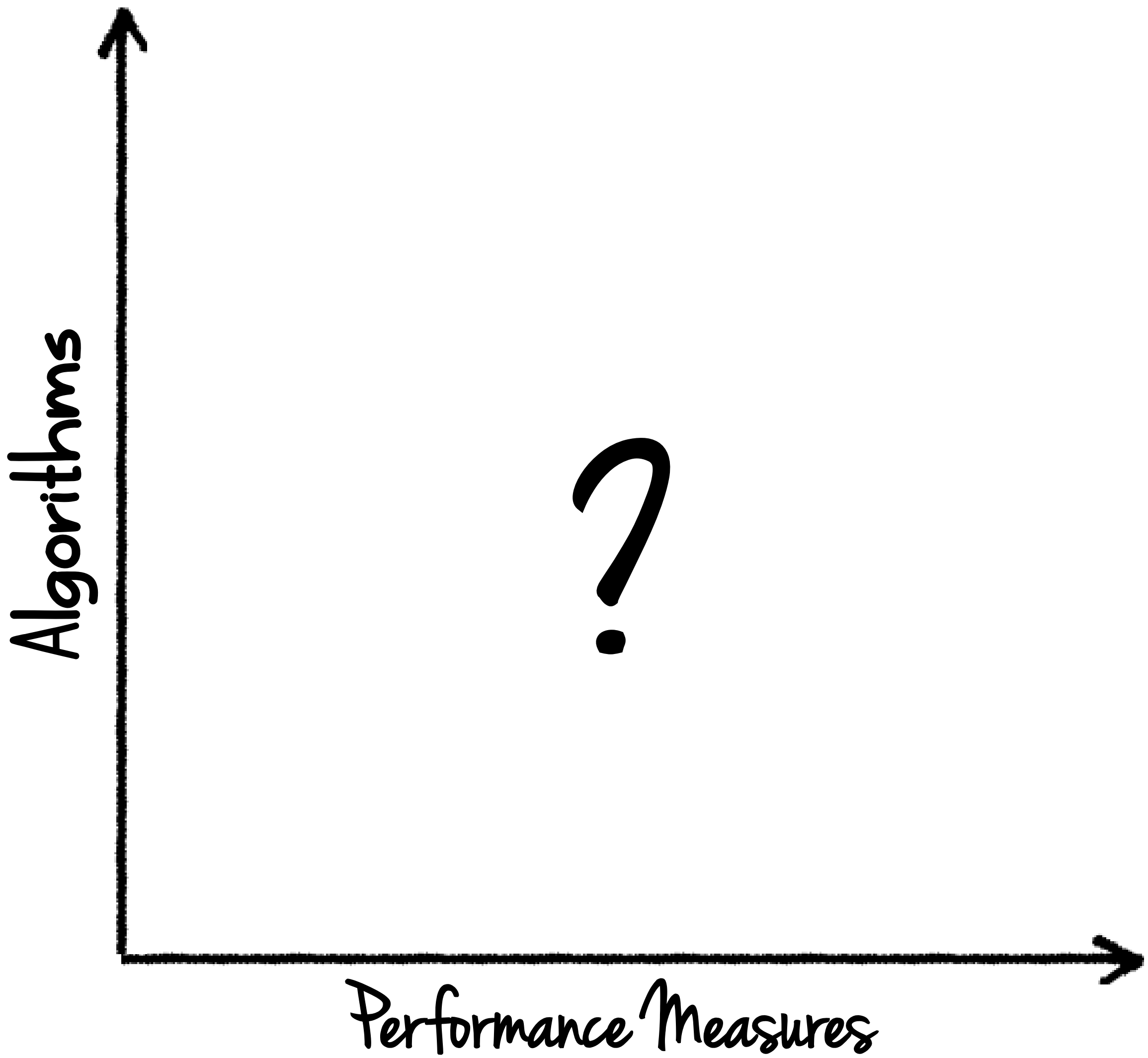
[PDF] Future directions in learning to rank
jmlr.csail.mit.edu/proceedings/papers/v14/.../chapelle11b.pdf
File Format: PDF/Adobe Acrobat - Quick View

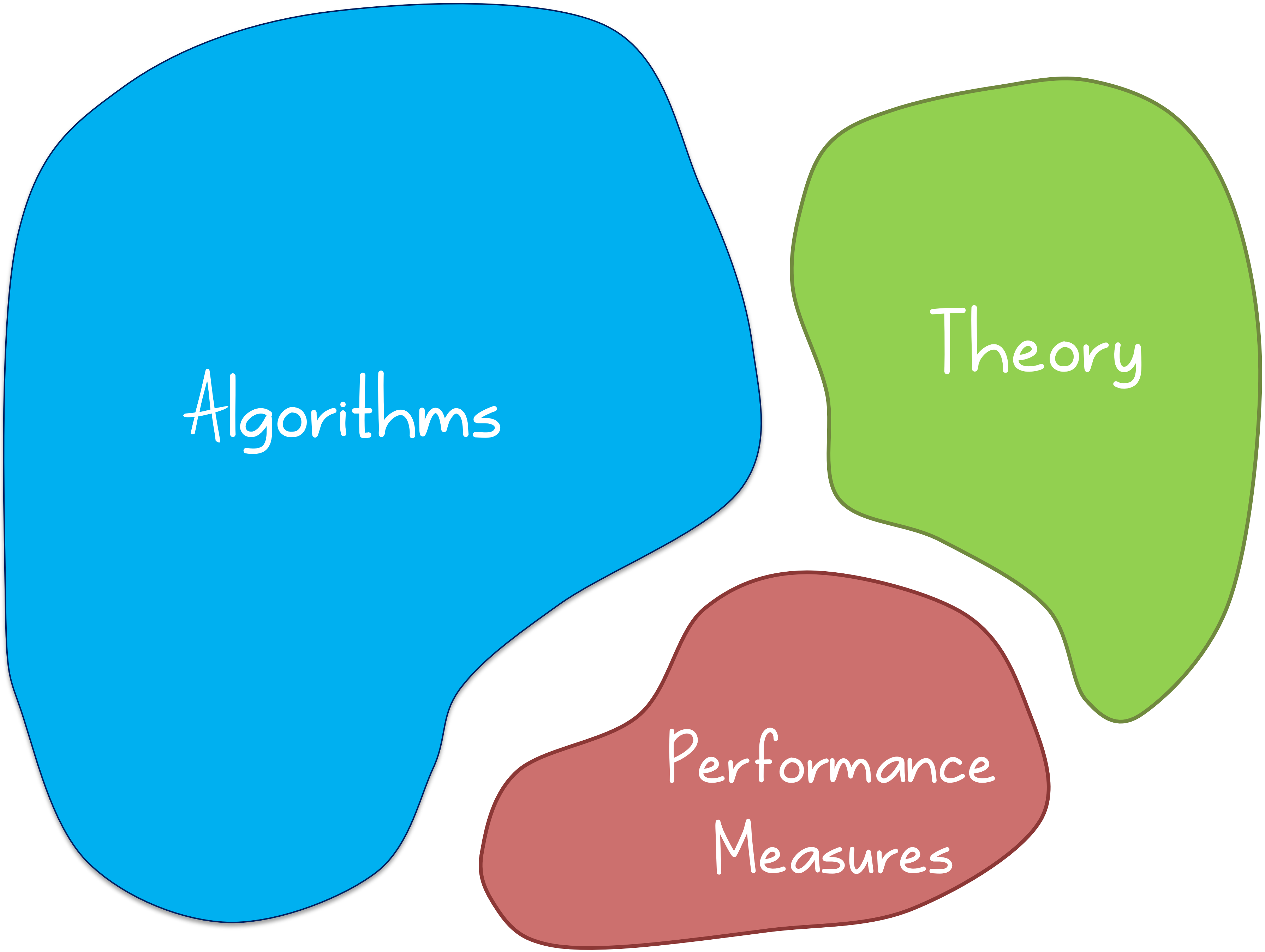
Average Precision

http://www.google.com/



http://www.tagxedo.com


Algorithms
?
Performance Measures


Algorithms
Theory
Performance Measures

Performance Measures

Binary Classification

AM-measure

G-mean

F-measure

...

Bipartite Ranking

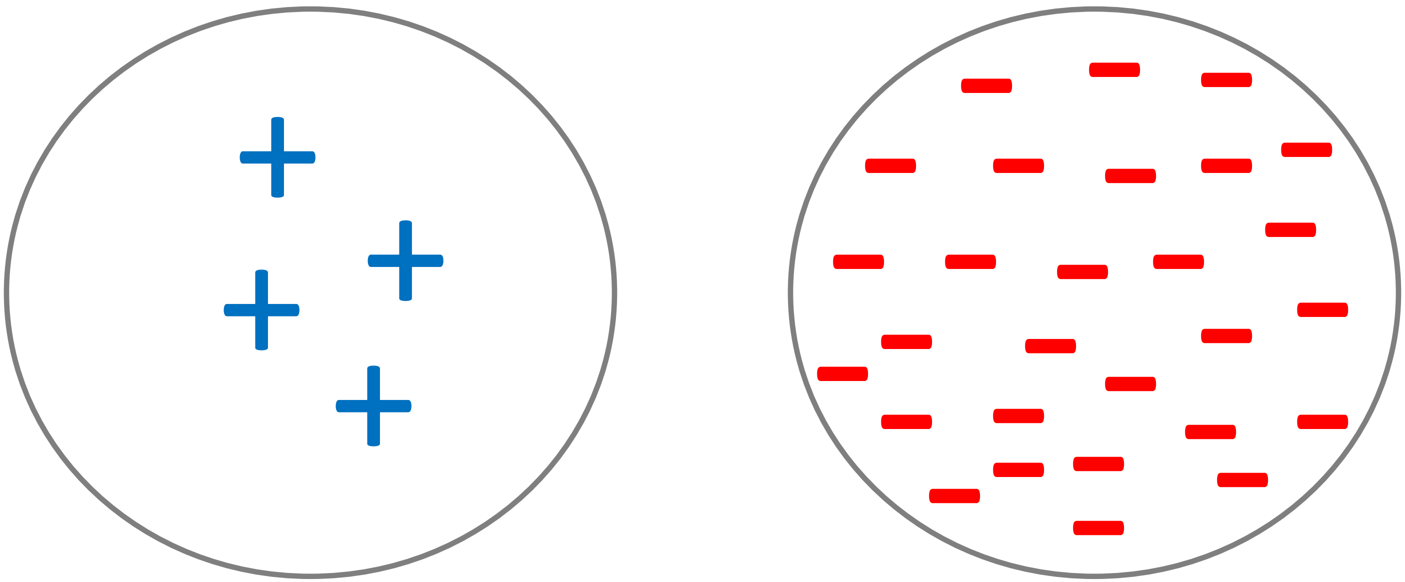
AUC

Precision@K

Average Precision

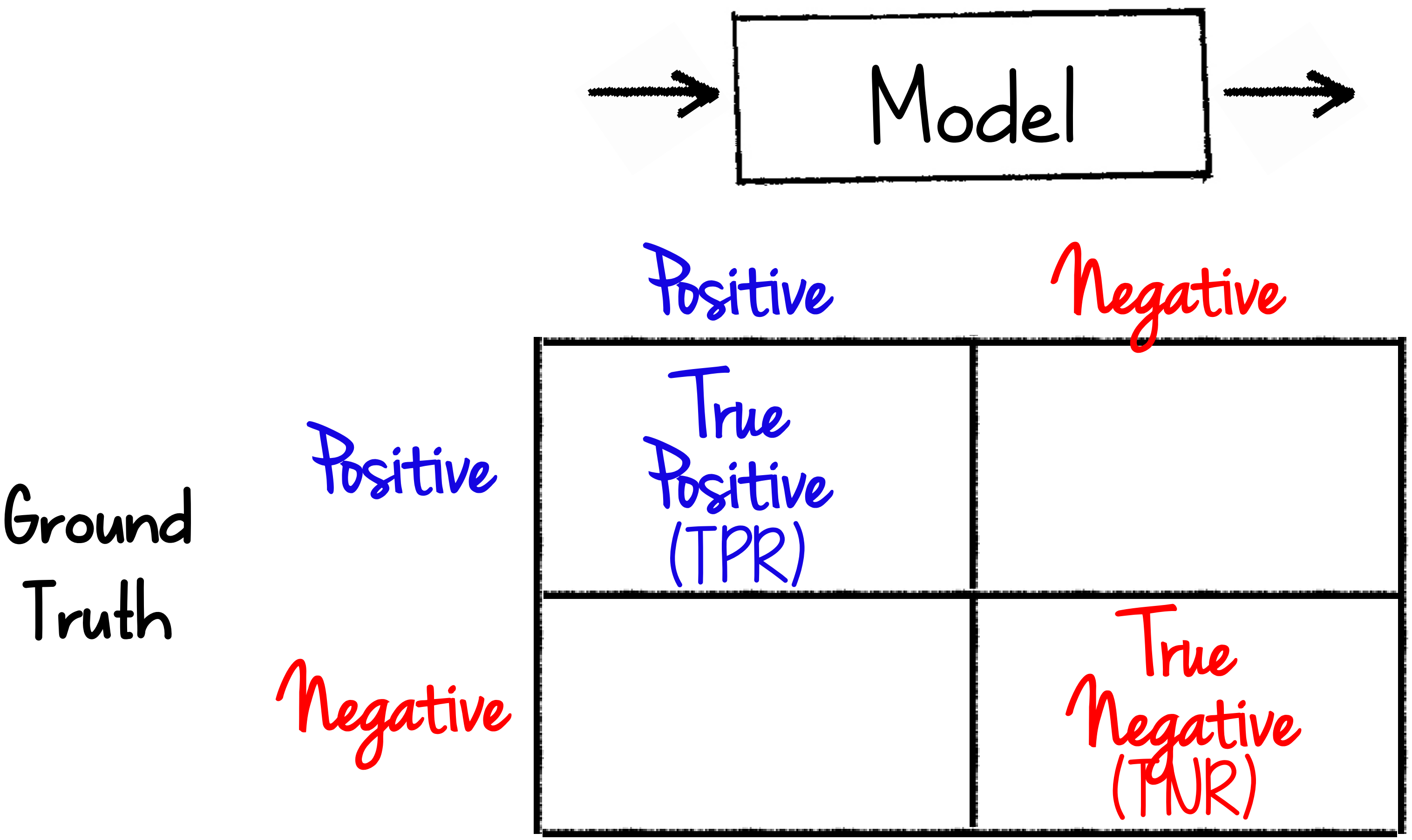
...

Class Imbalance



Standard Classification Error Ill-suited!

Class Imbalance



Class Imbalance

TPR TNR

Class Imbalance

AM-measure

$$\frac{\mathrm{TPR} + \mathrm{TNR}}{2}$$

Class Imbalance

G-mean

$$\sqrt{\mathrm{TPR} \times \mathrm{TNR}}$$

Class Imbalance

Recall

$$\cancel{\text{TPR}} = \frac{\text{no. of positive objects predicted as positive}}{\text{total no. of positive objects}}$$

$$\text{Precision} = \frac{\text{no. of positive objects predicted as positive}}{\text{total no. of objects predicted as positive}}$$

Class Imbalance

F-measure

$$\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Class Imbalance

Measure	Definition	References
A-Mean (AM)	$(\mathrm{TPR} + \mathrm{TNR})/2$	Chan & Stolfo (1998); Powers et al. (2005); Gu et al. (2009); KDD Cup 2001 challenge
G-Mean (GM)	$\sqrt{\mathrm{TPR} \cdot \mathrm{TNR}}$	Kubat & Matwin (1997); Daskalaki et al. (2006)
H-Mean (HM)	$2/(\frac{1}{\mathrm{TPR}} + \frac{1}{\mathrm{TNR}})$	Kennedy et al. (2009)
Q-Mean (QM)	$1 - ((\mathrm{FPR})^2 + (\mathrm{FNR})^2)/2$	Lawrence et al. (1998)
F_1	$2/(\frac{1}{\mathrm{Prec}} + \frac{1}{\mathrm{TPR}})$	Lewis & Gale (1994) Gu et al. (2009)
G-TP/PR	$\sqrt{\mathrm{TPR} \cdot \mathrm{Prec}}$	Daskalaki et al. (2006)

Bipartite Ranking

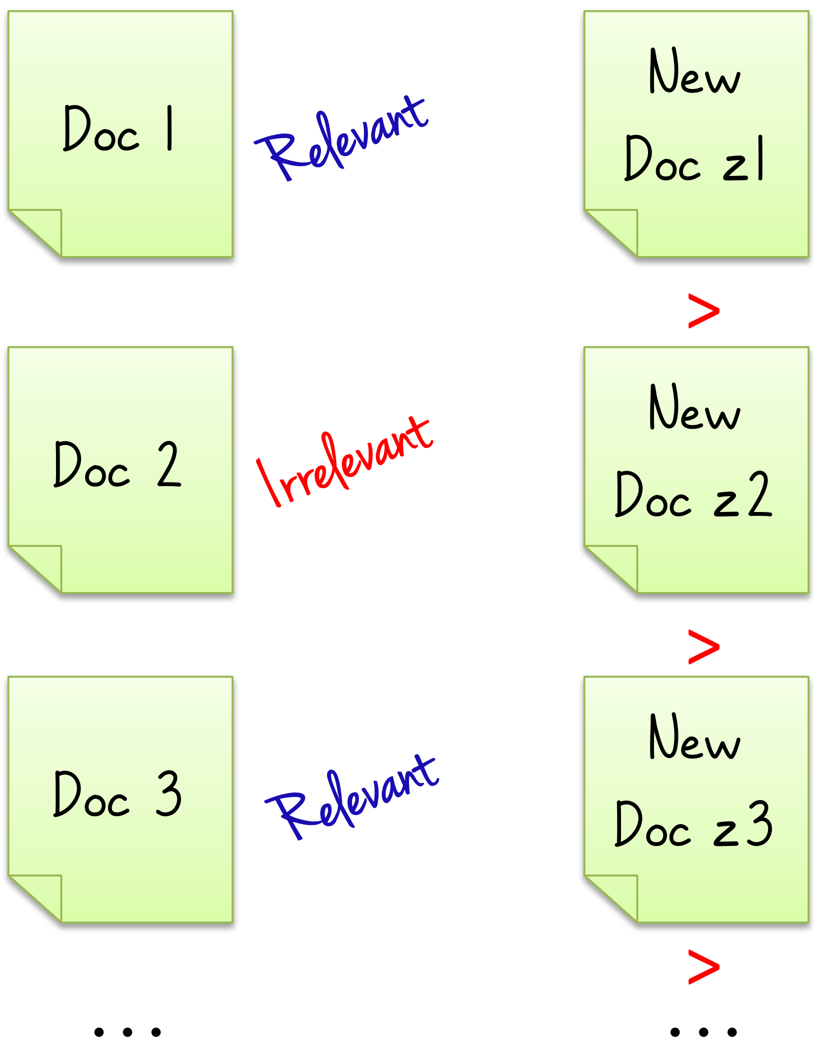
Doc 1 Relevant

Doc 2 Irrelevant

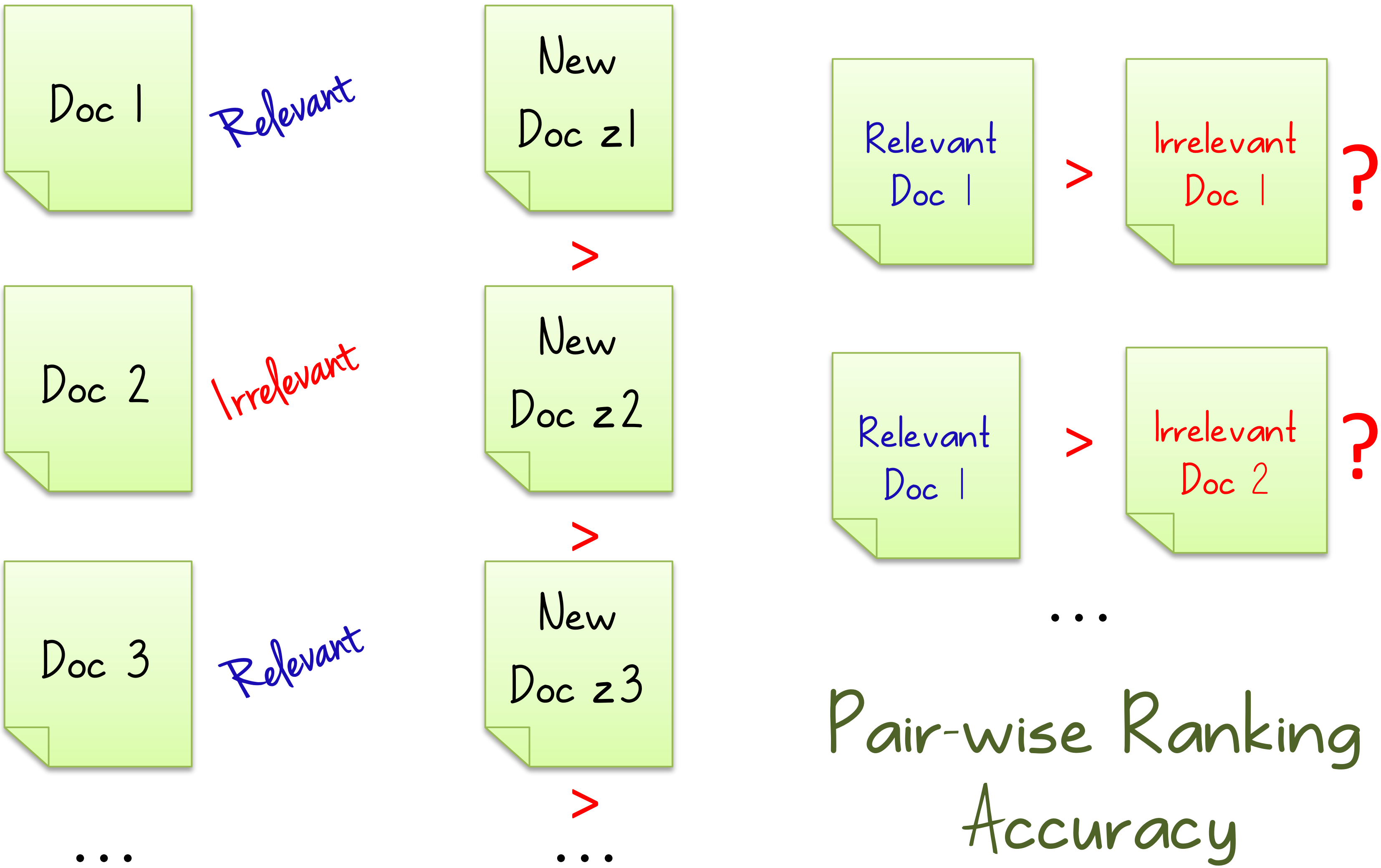
Doc 3 Relevant

...

Bipartite Ranking

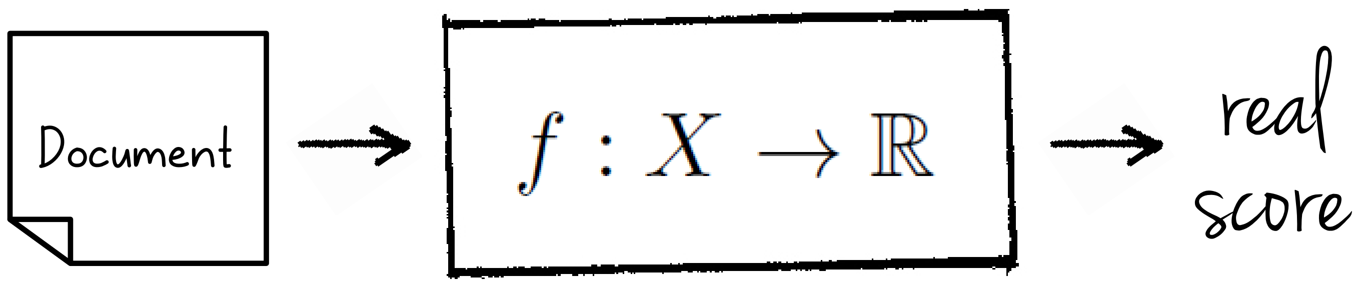


Bipartite Ranking



Bipartite Ranking

$$\{(x_1,y_1),(x_2,y_2),\ldots,(x_N,y_N)\}\subseteq\left(X\times\{\pm 1\}\right)^N$$



Bipartite Ranking

$$\{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\} \subseteq \left(X \times \{\pm 1\}\right)^N$$

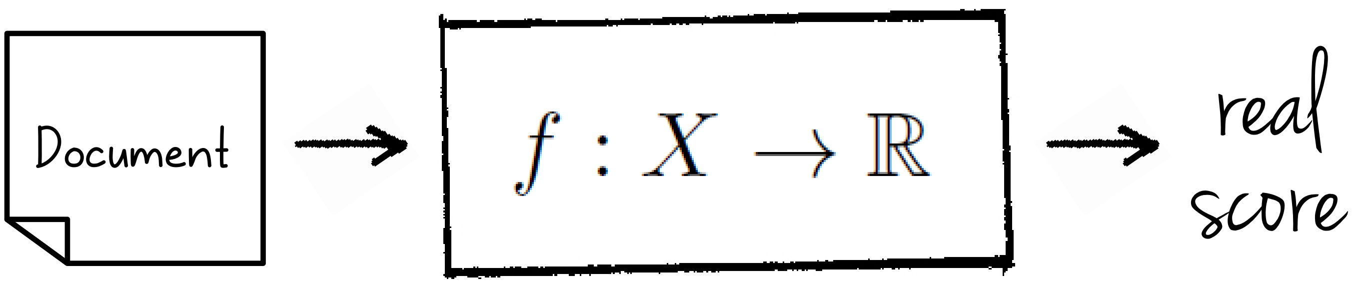
Document $\rightarrow$ $f : X \to \mathbb{R}$ $\rightarrow$ real score

Pair-wise Ranking
Accuracy

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \mathbf{1}\Big(\big(f(x_i) - f(x_j)\big)\left(y_i - y_j\right) > 0 \Big)$$

Bipartite Ranking

$$\{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\} \subseteq (X \times \{\pm 1\})^N$$

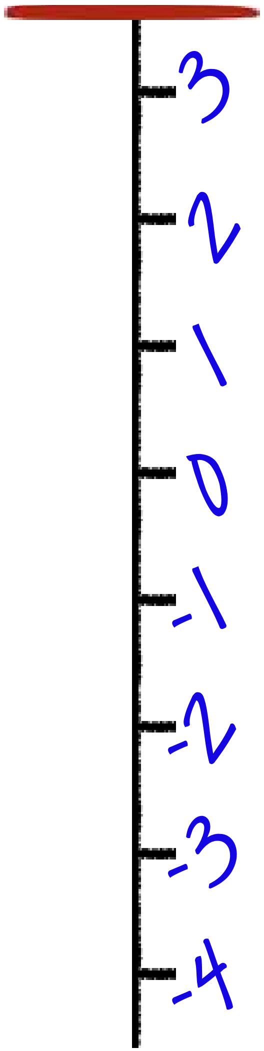


Pair-wise Ranking Accuracy

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \mathbf{1}\Big((f(x_i) - f(x_j))(y_i - y_j) > 0 \Big) + \frac{1}{2}\mathbf{1}((f(x_i) - f(x_j))(y_i - y_j) = 0)$$

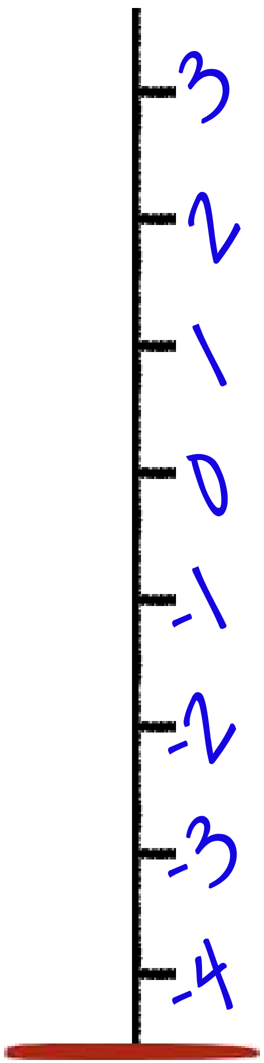
Bipartite Ranking

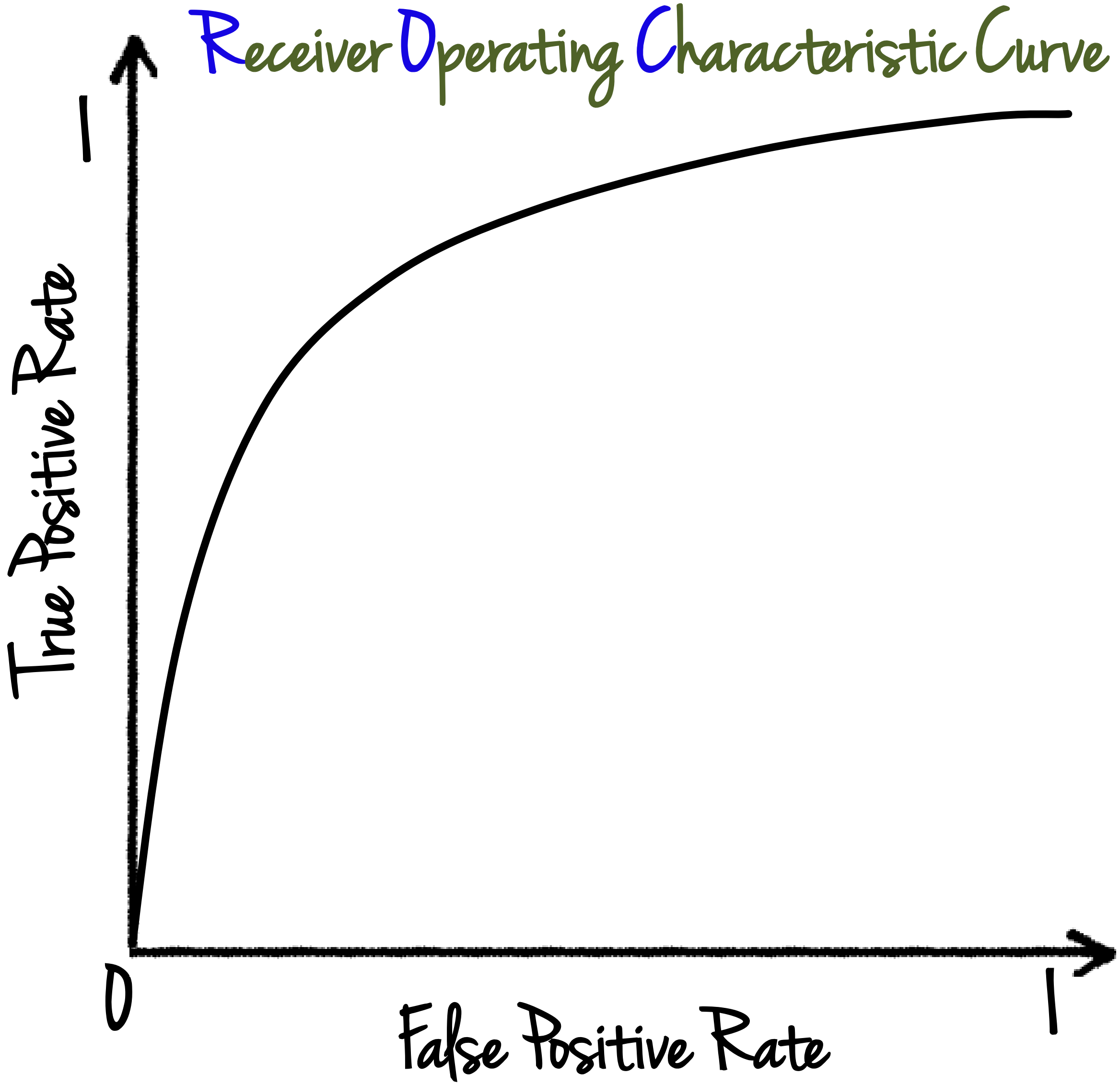
$f : X \to \mathbb{R}$



Bipartite Ranking

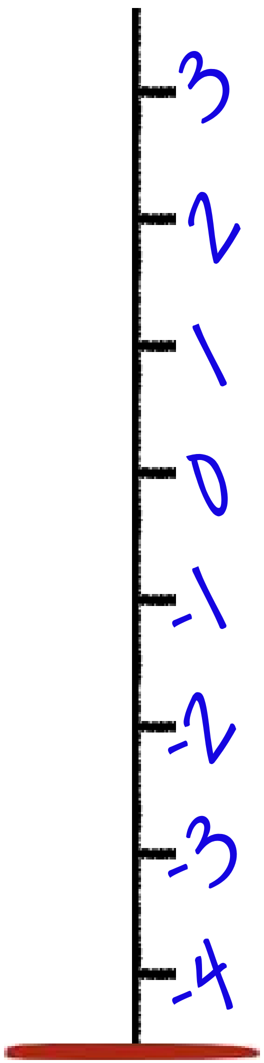
$f : X \to \mathbb{R}$

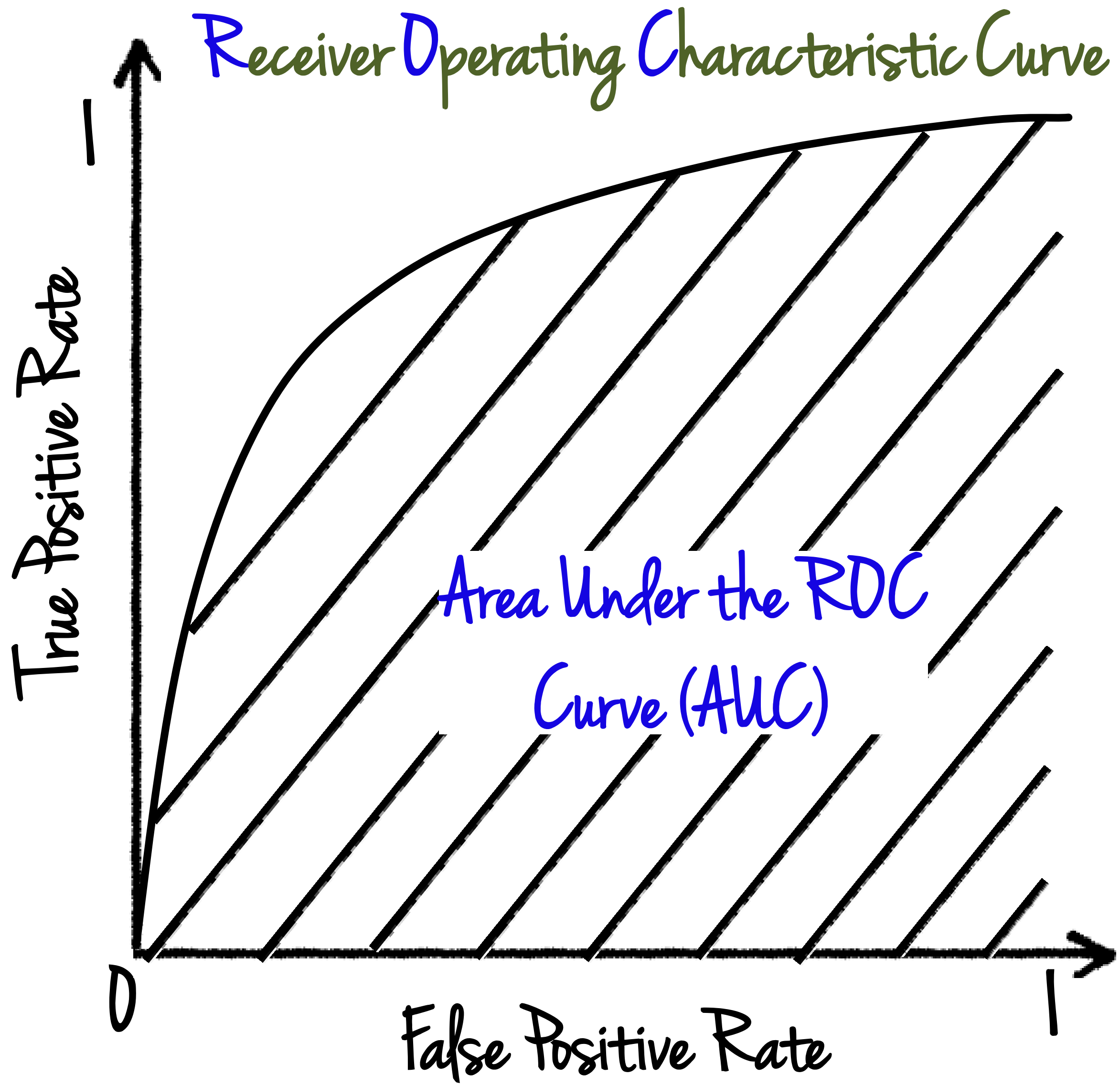




Bipartite Ranking

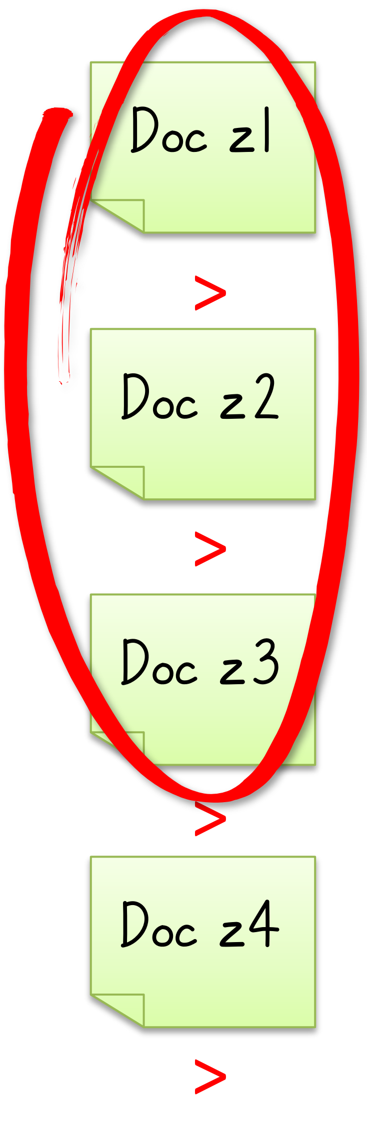
$f : X \to \mathbb{R}$





Bipartite Ranking

Bipartite Ranking

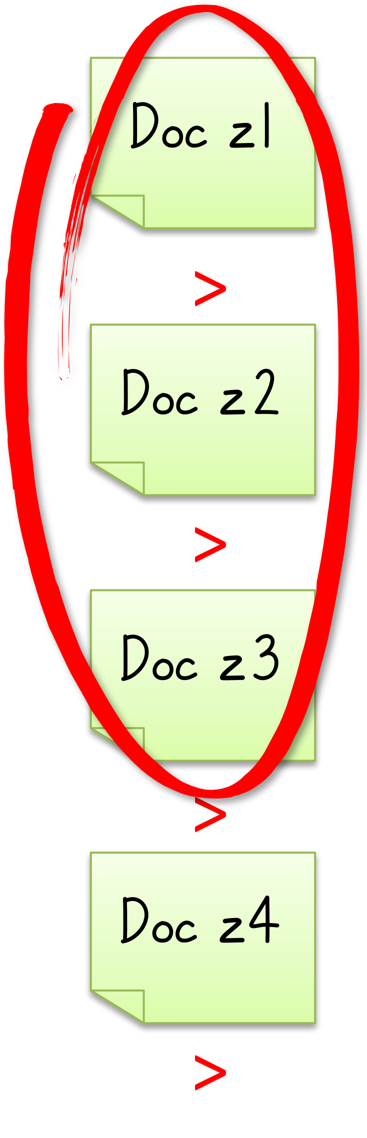


Positives in top K positions

Precision@K

$$\frac{1}{K}\sum_{i=1}^{K} \mathbf{1}\left(y_{(i)} = 1\right)$$

Bipartite Ranking



Positives in top K positions

Precision@K

$$\frac{1}{K}\sum_{i=1}^{K}\mathbf{1}\left(y_{(i)}=1\right)$$

Average Precision

$$\frac{1}{N}\sum_{K=1}^{N}\mathbf{1}\left(y_{(K)}=1\right)\text{Precision@}K$$

non-decomposable

G-mean

F-measure

H-mean

Precision@k

Average Precision

Q-mean

Discounted Cumulative Gain (DCG)

Mean Reciprocal Rank (MRR)

...

decomposable

0-1 Classification Error

Area Under the ROC Curve (AUC)

Cost-sensitive Error

AM-Measure

Performance Measures

Binary Classification

AM-measure

G-mean

F-measure

...

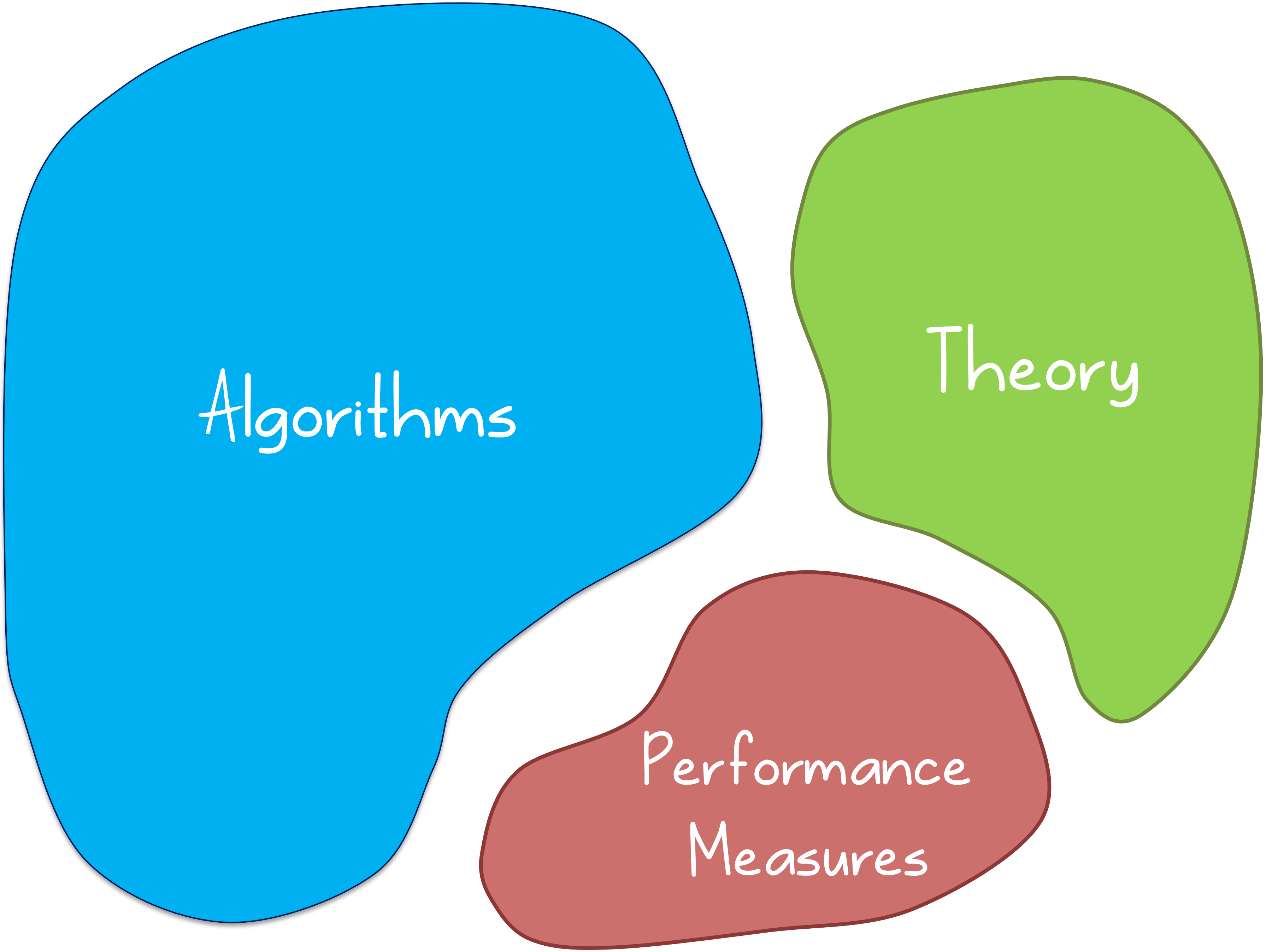
Bipartite Ranking

AUC

Precision@K

Average Precision

...


Algorithms
Theory
Performance Measures

Algorithms

Two Approaches

Plug-in

Risk Minimization

min {convex upper bound on loss}

Structural SVM

Algorithms

Two Approaches

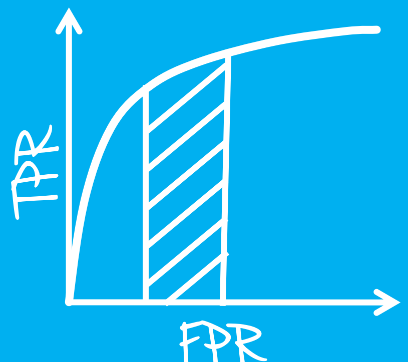
Plug-in

Risk Minimization

min {convex upper bound on loss}

Structural SVM

Case Study - Partial AUC



Two Families of Algorithms

$$S = \{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\} \subseteq \left(X \times \{\pm 1\}\right)^N$$

Goal:

max$\Big\{$ performance measure 'M' on S $\Big\}$

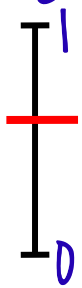
Two Families of Algorithms

$$S = \{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\} \subseteq (X \times \{\pm 1\})^N$$

Goal:

max { performance measure 'M' on S }

Plug-in



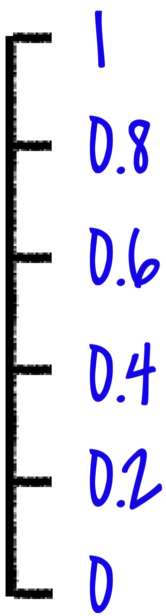
Risk Minimization

min { convex upper bound on loss }

Plug-in Approach

Learn a class probability estimator from S:

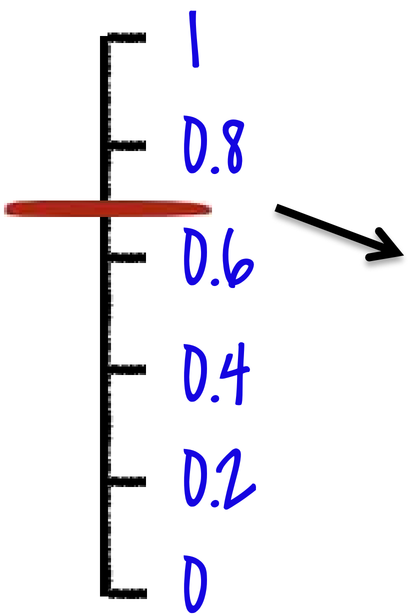
$$\widehat{\eta} : X \to [0, 1]$$



Plug-in Approach

Learn a class probability estimator from S:

$$\widehat{\eta} : X \to [0,1]$$

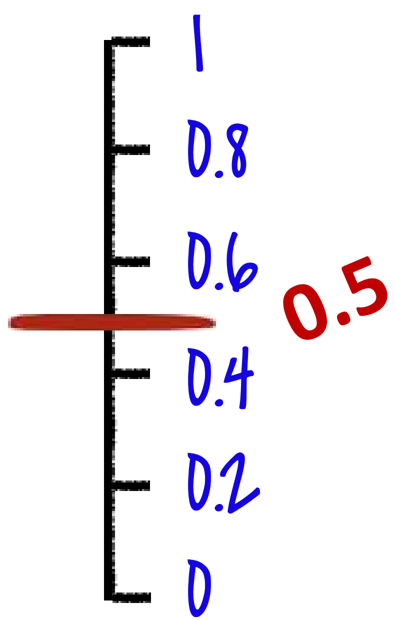


Use $\widehat{\eta}$ to construct a model. E.g.: choose threshold that optimizes performance measure 'M' on S

Plug-in Approach

Learn a class probability estimator from S:

$$\widehat{\eta} : X \to [0, 1]$$



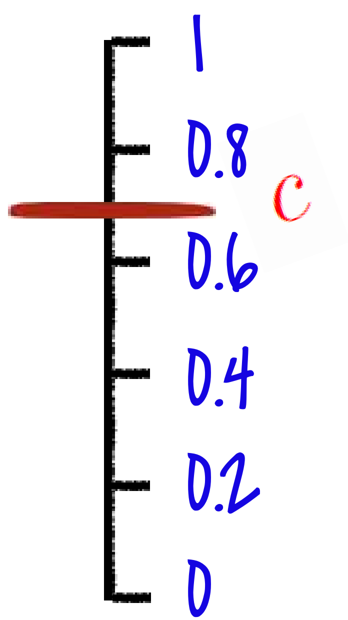
0-1 Classification Error

$$\sum_{i=1}^{N} \mathbf{1}\big(y_i \neq h(x_i)\big)$$

Plug-in Approach

Learn a class probability estimator from S:

$$\widehat{\eta} : X \to [0,1]$$



Cost-sensitive Classification Error

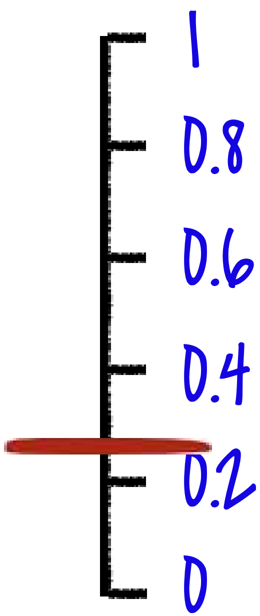
$$\sum_{i=1}^{N} c\mathbf{1}\big(y_i = 1, h(x_i) = -1\big) + (1-c)\mathbf{1}\big(y_i = -1, h(x_i) = 1\big)$$

$$c \in (0,1)$$

Plug-in Approach

Learn a class probability estimator from S:

$$\widehat{\eta} : X \to [0, 1]$$



Search over N+1 possible thresholds

F-measure

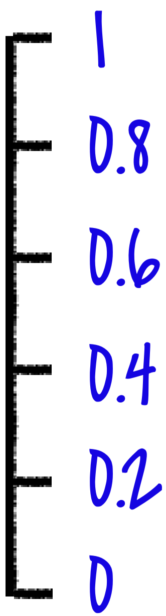
$$\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Yu, N., Chai, K.M.A., Lee, W.S., and Chieu, H.L.Optimizing F-Measures: A Tale of Two Approaches. In ICML 2012

Plug-in Approach

Learn a class probability estimator from S:

$$\widehat{\eta} : X \to [0,1]$$



Rank examples using $\widehat{\eta}$

AUC / Pair-wise Ranking Accuracy

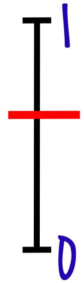
$$\sum_{i=1}^{N}\sum_{j=1}^{N} \mathbf{1}\Big(\big(f(x_i) - f(x_j)\big)\big(y_i - y_j\big) > 0\Big)$$

Two Families of Algorithms

$$S = \{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\} \subseteq \left(X \times \{\pm 1\}\right)^N$$

Goal:

max { performance measure 'M' on S }

Plug-in	Risk Minimization
	min { convex upper bound on loss }

Risk Minimization

~~max { performance measure 'M' on S }~~

min { convex surrogate objective on S }

Risk Minimization

~~max { performance measure 'M' on S }~~

min { convex surrogate objective on S }

Cost-sensitive Classification Error

$$\sum_{i=1}^{n} c\mathbf{1}\left(y_i = 1, h(x_i) = -1\right) + (1-c)\mathbf{1}\left(y_i = -1, h(x_i) = 1\right)$$

Risk Minimization

~~max { performance measure 'M' on S }~~

min { convex surrogate objective on S }

Cost-sensitive Classification Error

$$\cancel{\sum_{i=1}^{n} c\mathbf{1}(y_i = 1, h(x_i) = -1) + (1-c)\mathbf{1}(y_i = -1, h(x_i) = 1)}$$

$$\sum_{i=1}^{n} c\mathbf{1}(y_i = 1)\ell(1, f(x_i)) + (1-c)\mathbf{1}(y_i = -1)\ell(-1, f(x_i))$$

convex loss $\ell : \{\pm 1\} \times \mathbb{R} \to \mathbb{R}_+$

Risk Minimization

~~max { performance measure 'M' on S }~~

min { convex surrogate objective on S }

AUC / Pair-wise Ranking Accuracy

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{1}\Big(\big(f(x_i)-f(x_j)\big)\big(y_i-y_j\big)>0\Big)$$

Risk Minimization

~~max { performance measure 'M' on S }~~

min { convex surrogate objective on S }

AUC / Pair-wise Ranking Accuracy

$$\cancel{\sum_{i=1}^{N}\sum_{j=1}^{N} \mathbf{1}\Big(\big(f(x_i) - f(x_j)\big)\big(y_i - y_j\big) > 0\Big)}$$

$$\sum_{i=1}^{N}\sum_{j=1}^{N} \ell\big(f(x_i) - f(x_j)\big)\mathbf{1}\big((y_i - y_j) > 0\big)$$

convex loss $\ell : \mathbb{R} \to \mathbb{R}_+$

G-mean F-measure H-mean

Precision@k Average Precision Q-mean

Discounted Cumulative Gain (DCG) Mean Reciprocal Rank (MRR)

...

decomposable

0-1 Classification Error

Area Under the ROC Curve (AUC)

Cost-sensitive Error

AM-Measure

non-decomposable

G-mean

F-measure

H-mean

Precision@k

Average Precision

Q-mean

Discounted Cumulative Gain (DCG)

Mean Reciprocal Rank (MRR)

...

decomposable

0-1 Classification Error

Area Under the ROC Curve (AUC)

Cost-sensitive Error

AM-Measure

Risk Minimization

~~max { performance measure 'M' on S }~~

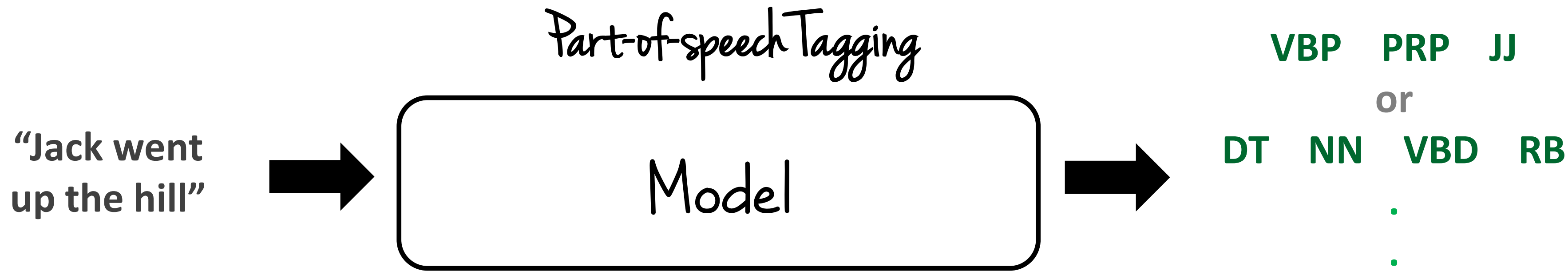
min { convex surrogate objective on S }

Structural SVM

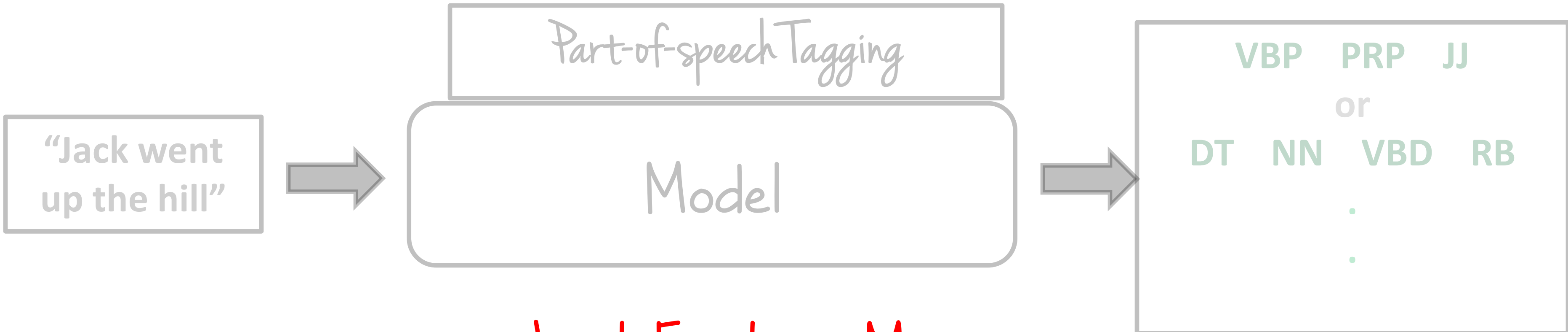
F-measure PRBEP Precision@k Average Precision DCG MRR AUC ...

Joachims, T. A Support Vector Method for Multivariate Performance Measures. In ICML, 2005.

Structural SVM for Multivariate Performance Measures



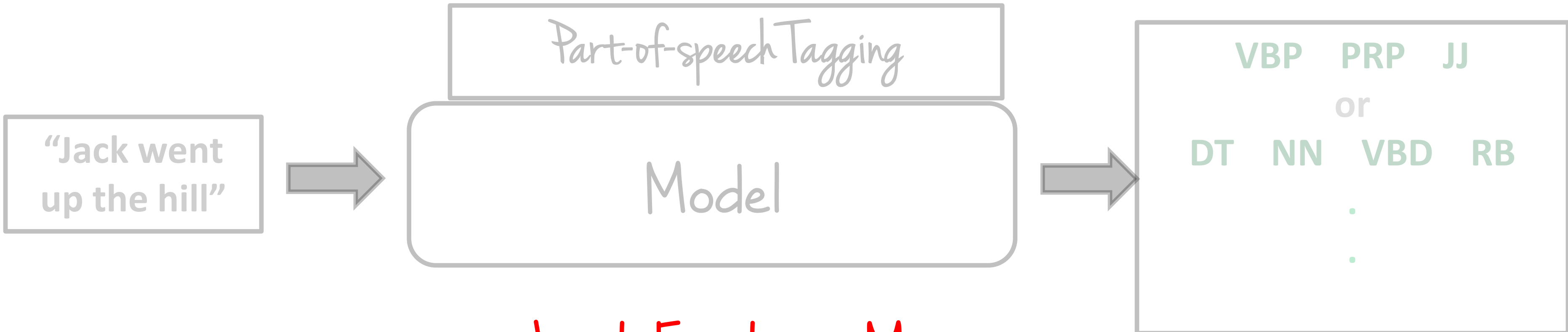
Structural SVM for Multivariate Performance Measures



Joint Feature Map

$$\phi(\text{input}, \text{output}) \longmapsto \mathbb{R}^d$$

Structural SVM for Multivariate Performance Measures

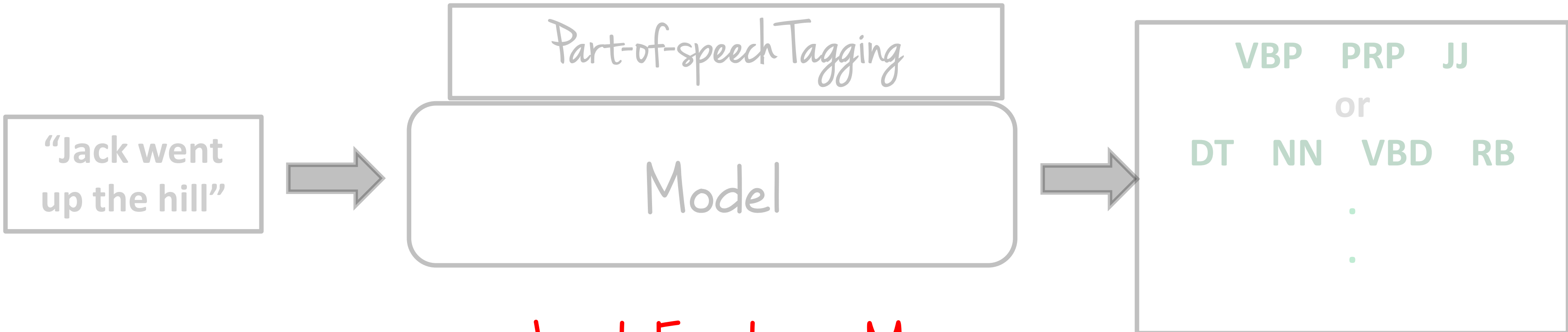


Joint Feature Map

$$w^\top \phi(\text{input}, \text{output}) \longmapsto \mathbb{R}$$

Discriminant Function

Structural SVM for Multivariate Performance Measures



Joint Feature Map

Discriminant Function

$$w^\top \phi(\text{input}, \text{output}) \longmapsto \mathbb{R}$$

$$x \rightarrow \boxed{\operatorname*{argmax}_{\pi} \; w^\top \phi(x, \pi)} \rightarrow \pi$$

Maximum over an exponentially large set

Structural SVM for Multivariate Performance Measures

Input: Training set S

Structural SVM for Multivariate Performance Measures

Input: Training set S

Ingredient 1

Output Space: Π

Structural SVM for Multivariate Performance Measures

Input: Training set S

Ingredient 1

Output Space: Π

Ingredient 2

$$\phi : (\underbrace{X \times \{\pm 1\}}_{S})^N \times \Pi \to \mathbb{R}^d$$

Structural SVM for Multivariate Performance Measures

Input: Training set S

Ingredient 1

Output Space: Π

Ingredient 2

$$\phi : (\underbrace{X \times \{\pm 1\}}_{S})^N \times \Pi \rightarrow \mathbb{R}^d$$

Ingredient 3

$$\Delta : \Pi \times \Pi \rightarrow \mathbb{R}_+$$

Structural SVM for Multivariate Performance Measures

E.g.: F-measure Optimization

$$X \subseteq \mathbb{R}^d$$

Structural SVM for Multivariate Performance Measures

E.g.: F-measure Optimization

$$X \subseteq \mathbb{R}^d$$

Ingredient 1

Output Space: $\Pi = \{\pm 1\}^N$

Structural SVM for Multivariate Performance Measures

E.g.: F-measure Optimization

$$X \subseteq \mathbb{R}^d$$

Ideal

$$\pi^* = [y_1, \ldots, y_N]$$

Ingredient I

Output Space: $\Pi = \{\pm 1\}^N$

Structural SVM for Multivariate Performance Measures

E.g.: F-measure Optimization

$$X \subseteq \mathbb{R}^d$$

Ideal

$$\pi^* = [y_1, \ldots, y_N]$$

Ingredient 1

Output Space: $\Pi = \{\pm 1\}^N$

Ingredient 2

$$\phi(S, \pi) = \sum_{i=1}^{N} \pi_i x_i$$

Structural SVM for Multivariate Performance Measures

E.g.: F-measure Optimization

$$X \subseteq \mathbb{R}^d$$

Ideal

$$\pi^* = [y_1, \ldots, y_N]$$

Ingredient 1

Output Space: $\Pi = \{\pm 1\}^N$

Ingredient 2

$$\phi(S,\pi) = \sum_{i=1}^{N} \pi_i x_i$$

$$\boxed{\begin{aligned} &\operatorname*{argmax}_{\pi \in \Pi} w^\top \phi(S,\pi) \\ &= \operatorname*{argmax}_{\pi \in \Pi} \sum_{i=1}^{N} \pi_i (w^\top x_i) \end{aligned}}$$

Structural SVM for Multivariate Performance Measures

E.g.: F-measure Optimization

$$X \subseteq \mathbb{R}^d$$

Ideal

$$\pi^* = [y_1, \ldots, y_N]$$

Ingredient 1

Output Space: $\Pi = \{\pm 1\}^N$

Ingredient 2

$$\phi(S, \pi) = \sum_{i=1}^{N} \pi_i x_i$$

Ingredient 3

$$\Delta(\pi, \pi^*) = 1 - \text{F-measure}(\pi^*, \pi)$$

Structural SVM for Multivariate Performance Measures

$$\min_{w,\ \xi\geq 0} \ \frac{1}{2}||w||^2 \ + \ C\xi$$

s.t.

$$\forall \pi \in \Pi : \quad w^\top \big(\phi(S,\ \pi^*) - \phi(S,\ \pi)\big) \ \geq \ \Delta(\pi, \pi^*) - \xi$$

Structural SVM for Multivariate Performance Measures

$$\min_{w,\ \xi \geq 0} \ \frac{1}{2}||w||^2 \ + \ C\xi$$

s.t.

$$\boxed{\forall \pi \in \Pi :} \ \ w^\top \left(\phi(S,\ \pi^*) - \phi(S,\ \pi)\right) \ \geq \ \Delta(\pi, \pi^*) - \xi$$

Exponential Number
of Outputs

Structural SVM for Multivariate Performance Measures

Upper Bound on loss Δ

$$\min_{w,\ \xi \geq 0} \ \frac{1}{2}||w||^2 \ + \ C\xi$$

s.t.

$$\boxed{\forall \pi \in \Pi :} \ w^\top \left(\phi(S,\ \pi^*) - \phi(S,\ \pi)\right) \ \geq \Delta(\pi, \pi^*) - \xi$$

Exponential Number of Outputs

Structural SVM for Multivariate Performance Measures

Cutting-plane Procedure

$$\min_{w,\ \xi \geq 0} \ \frac{1}{2}||w||^2 \ + \ C\xi$$

s.t.

$$\forall \pi \in \Pi : \quad w^\top \big(\phi(S,\ \pi^*) - \phi(S,\ \pi)\big) \ \geq \ \Delta(\pi, \pi^*) - \xi$$

Repeat:

1. Solve OP for a subset of constraints.

Structural SVM for Multivariate Performance Measures

Cutting-plane Procedure

$$\min_{w,\ \xi \geq 0} \frac{1}{2}||w||^2 + C\xi$$

s.t.

$$\forall \pi \in \Pi : \boxed{w^\top \left(\phi(S,\ \pi^*) - \phi(S,\ \pi)\right) \ \geq \ \Delta(\pi, \pi^*) - \xi}$$

Repeat:

1. Solve OP for a subset of constraints.
2. Add the most violated constraint.

Structural SVM for Multivariate Performance Measures

Cutting-plane Procedure

Converges in constant number of iterations

Repeat:

$$\min_{w,\ \xi \geq 0} \frac{1}{2}||w||^2 + C\xi$$

s.t.

$$\forall \pi \in \Pi : \boxed{w^\top \left(\phi(S,\ \pi^*) - \phi(S,\ \pi)\right) \ \geq\ \Delta(\pi, \pi^*) - \xi}$$

1. Solve OP for a subset of constraints.
2. Add the most violated constraint.

Structural SVM for Multivariate Performance Measures

Performance Measure	Run-time	References
F-measure PBREP Precision@K	$O(N^2)$	Joachims (2005);
AUC Partial AUC Pos@Top	$O(N \log N)$	Joachims (2005, 2006); Narasimhan & Agarwal (2013 a,b)
Average Precision	$O(N^2)$	Yue et al. (2007)
Discounted Cumulative Gain (DCG) *	$O(Nk)$	Chakrabarti et al. (2008)
Mean Reciprocal Rank (MRR) *	$O(N \log N + k^2)$	Chakrabarti et al. (2008)

* performance measure truncated to top ‘k’ positions

Algorithms

Two Approaches

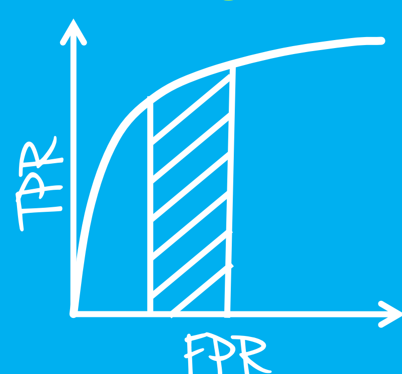
Plug-in

Risk Minimization

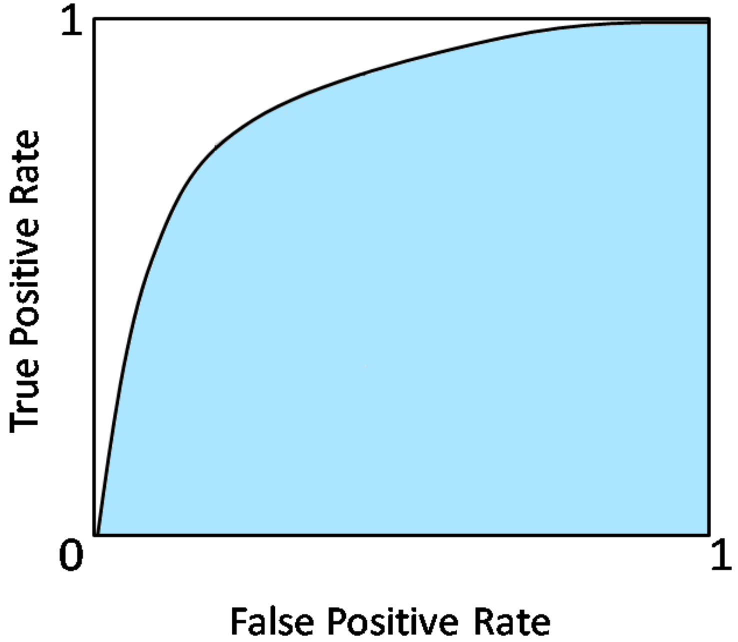
min {convex upper bound on loss}

Structural SVM

Case Study - Partial AUC

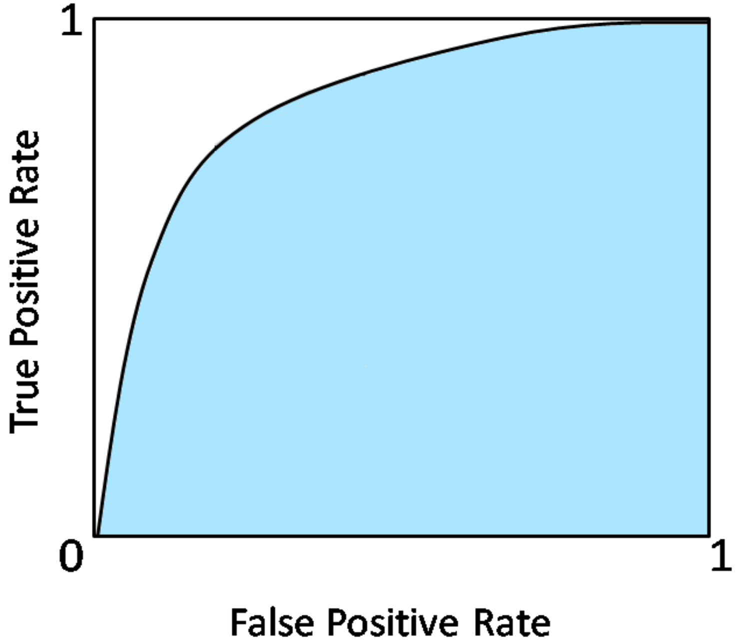


Partial AUC?



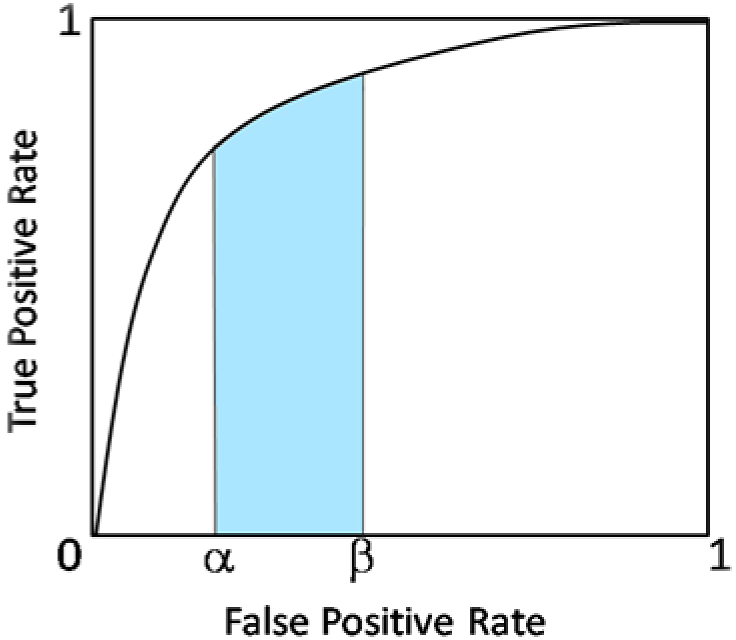
Full AUC

Partial AUC?



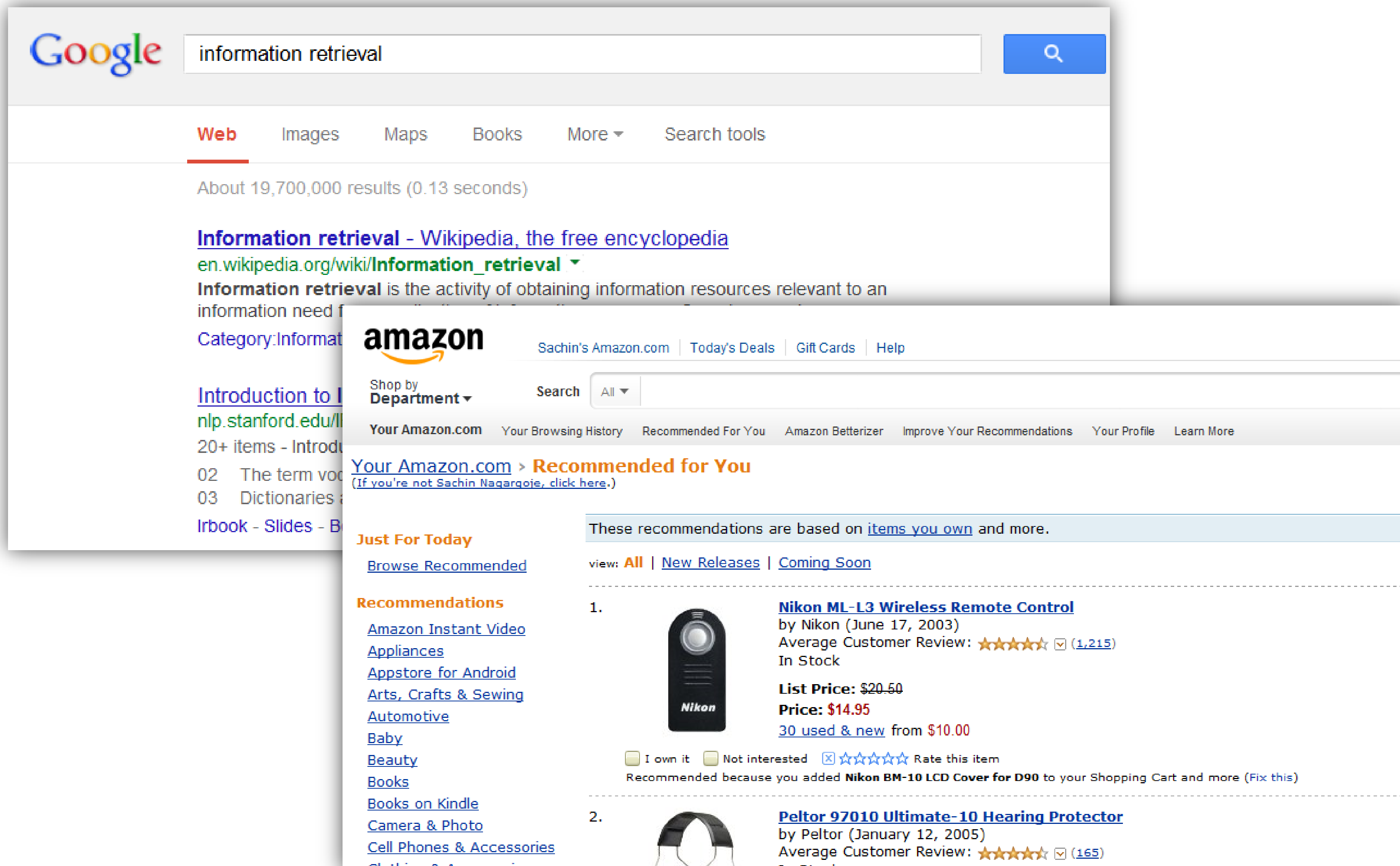
Full AUC

Vs



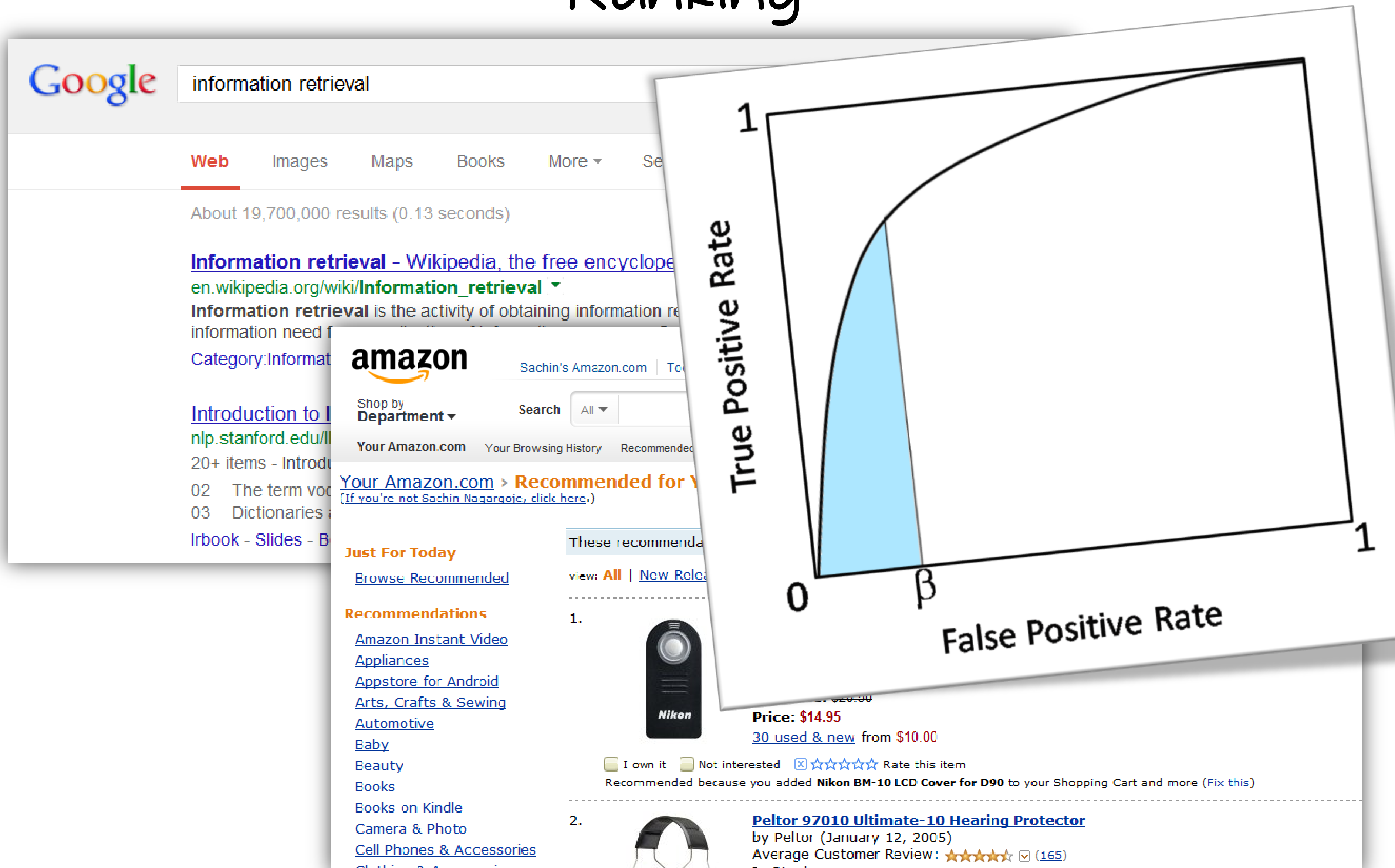
Partial AUC

Ranking



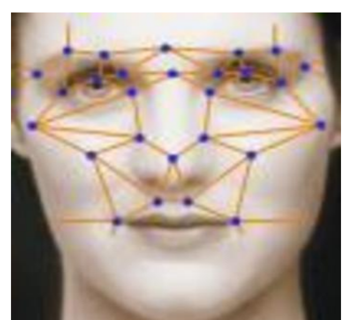
Google
information retrieval
Web
Images
Maps
Books
More
Search tools
About 19,700,000 results (0.13 seconds)
Information retrieval - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Information_retrieval
Information retrieval is the activity of obtaining information resources relevant to an
information need f
Category:Informat
Introduction to I
nlp.stanford.edu/I
20+ items - Introdu
02 The term voc
03 Dictionaries
Irbook - Slides - B
amazon
Sachin's Amazon.com
Today's Deals
Gift Cards
Help
Shop by
Department
Search
All
Your Amazon.com
Your Browsing History
Recommended For You
Amazon Betterizer
Improve Your Recommendations
Your Profile
Learn More
Your Amazon.com › Recommended for You
(If you're not Sachin Nagargoje, click here.)
Just For Today
Browse Recommended
Recommendations
Amazon Instant Video
Appliances
Appstore for Android
Arts, Crafts & Sewing
Automotive
Baby
Beauty
Books
Books on Kindle
Camera & Photo
Cell Phones & Accessories
Clothing & Accessories
These recommendations are based on items you own and more.
view: All | New Releases | Coming Soon
1.
Nikon ML-L3 Wireless Remote Control
by Nikon (June 17, 2003)
Average Customer Review: (1,215)
In Stock
List Price: $20.50
Price: $14.95
30 used & new from $10.00
Nikon
I own it
Not interested
Rate this item
Recommended because you added Nikon BM-10 LCD Cover for D90 to your Shopping Cart and more (Fix this)
2.
Peltor 97010 Ultimate-10 Hearing Protector
by Peltor (January 12, 2005)
Average Customer Review: (165)

Ranking



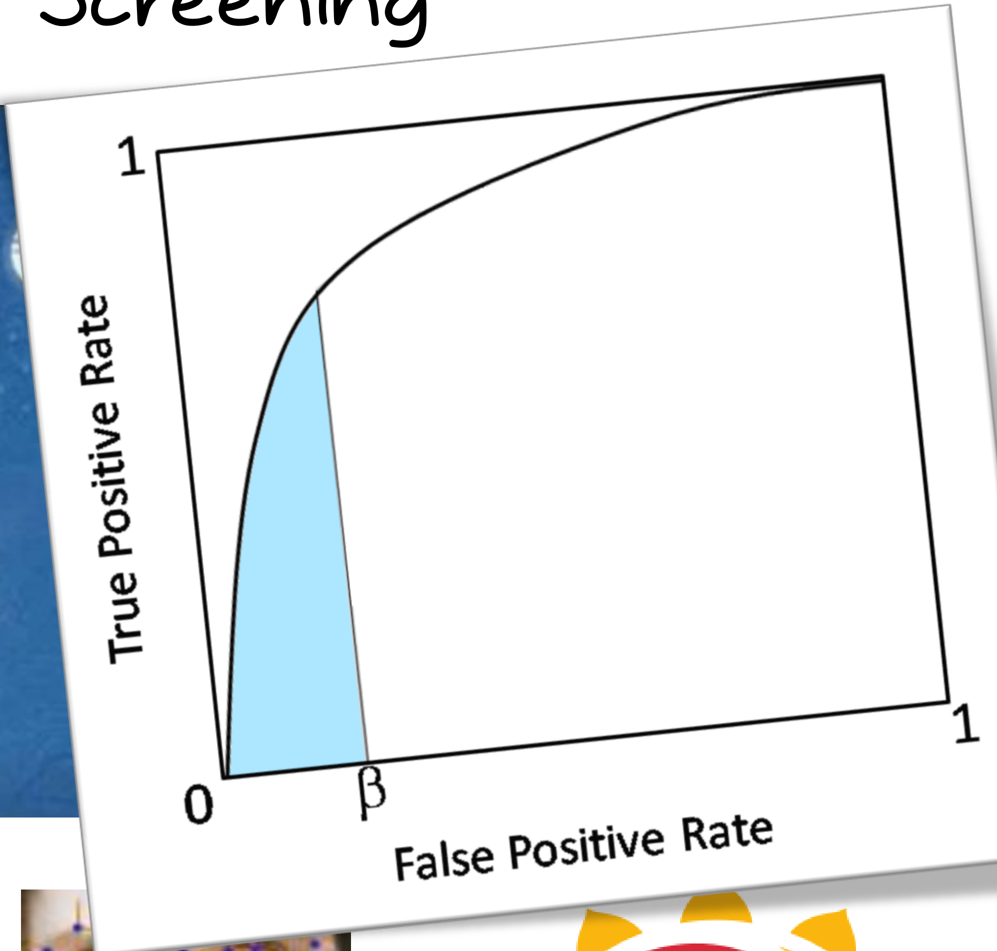
Google
information retrieval
Web
Images
Maps
Books
More
About 19,700,000 results (0.13 seconds)
Information retrieval - Wikipedia, the free encyclope
en.wikipedia.org/wiki/Information_retrieval
Information retrieval is the activity of obtaining information re
Introduction to I
nlp.stanford.edu/I
amazon
Sachin's Amazon.com
Shop by Department
Search
All
Your Amazon.com
Your Browsing History
Your Amazon.com › Recommended for
(If you're not Sachin Nagarojie, click here.)
Just For Today
Browse Recommended
Recommendations
Amazon Instant Video
Appliances
Appstore for Android
Arts, Crafts & Sewing
Automotive
Baby
Beauty
Books
Books on Kindle
Camera & Photo
Cell Phones & Accessories
Clothing & Accessories
These recommenda
view: All | New Relea
1.
Nikon
Price: $14.95
30 used & new from $10.00
I own it
Not interested
Rate this item
Recommended because you added Nikon BM-10 LCD Cover for D90 to your Shopping Cart and more (Fix this)
2.
Peltor 97010 Ultimate-10 Hearing Protector
by Peltor (January 12, 2005)
Average Customer Review: (165)
In Stock
1
True Positive Rate
0
β
1
False Positive Rate

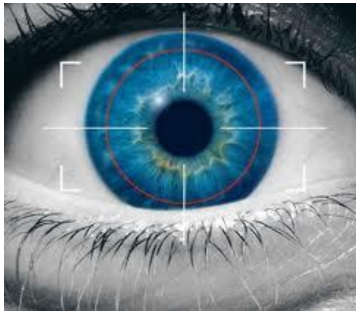
Biometric Screening

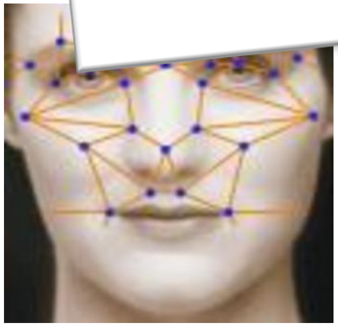


AADHAAR

Biometric Screening

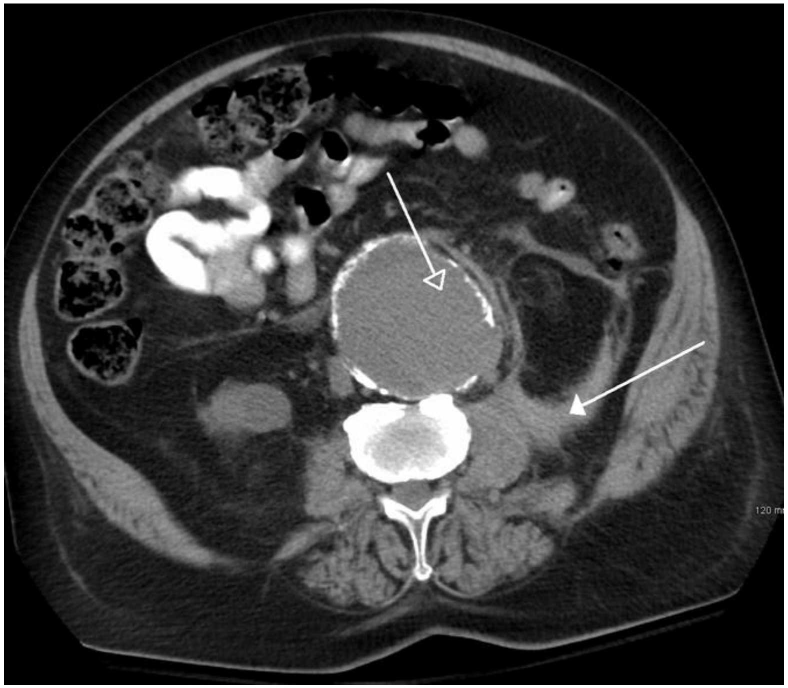


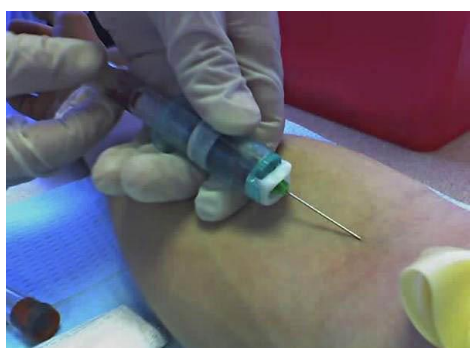




AADHAAR

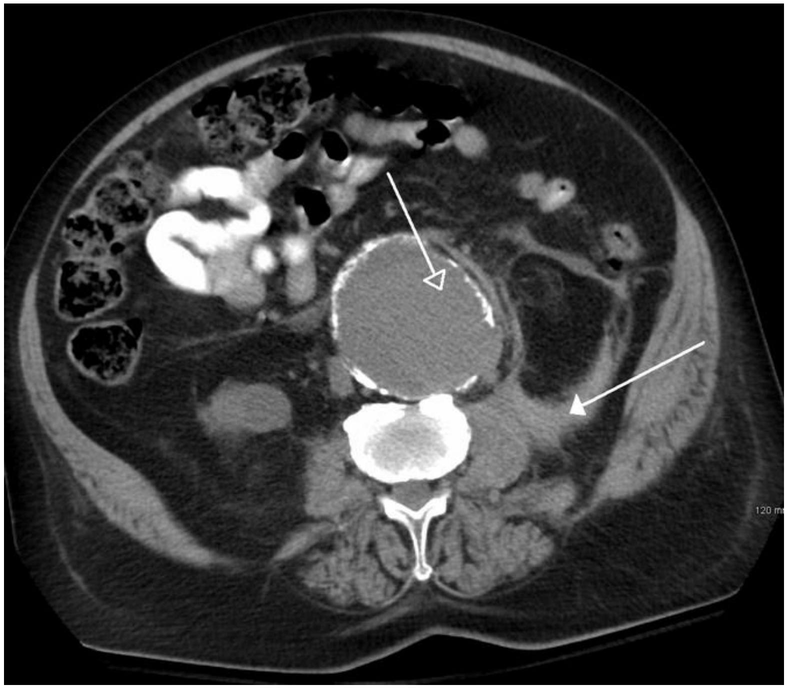
Medical Diagnosis

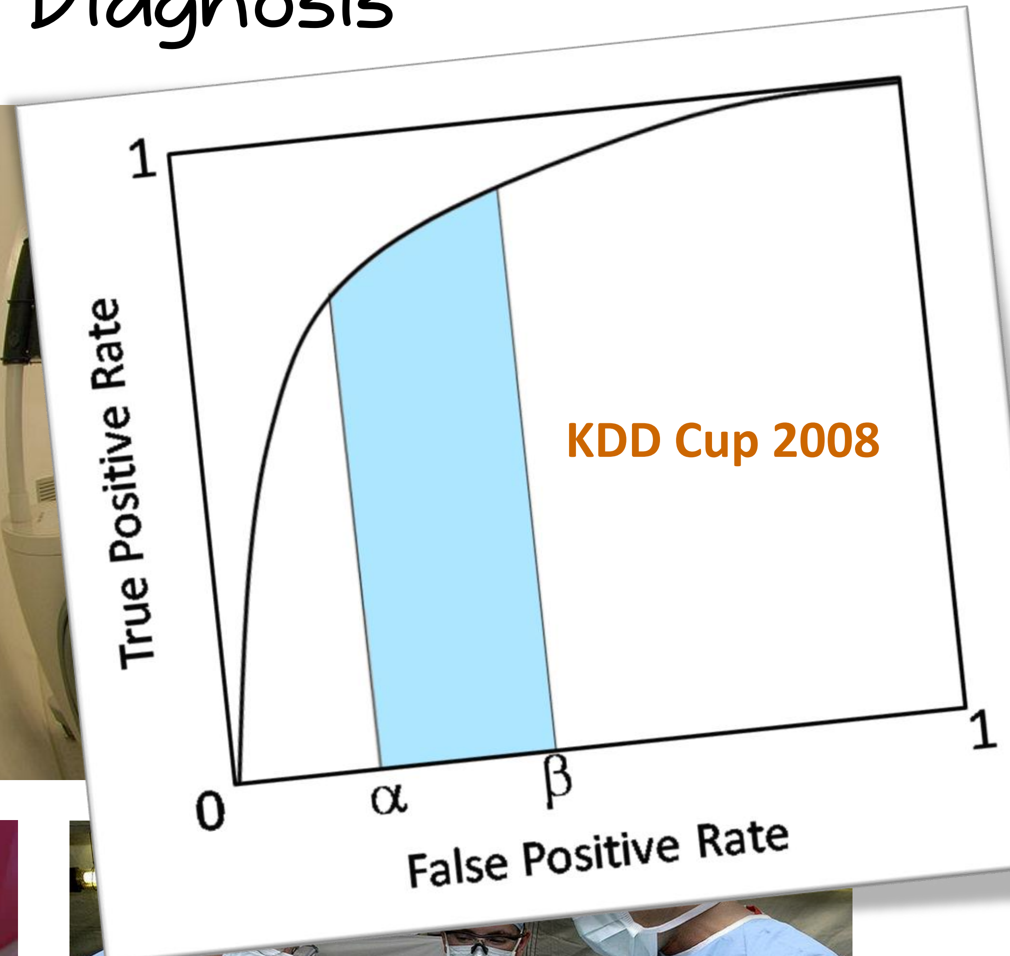


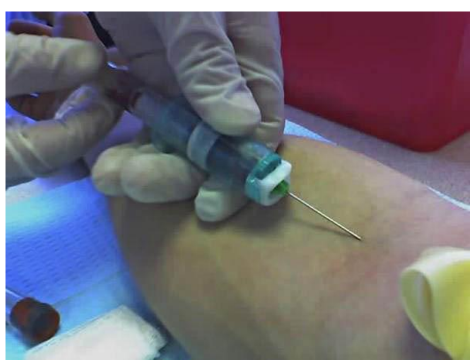


http://en.wikipedia.org/

Medical Diagnosis



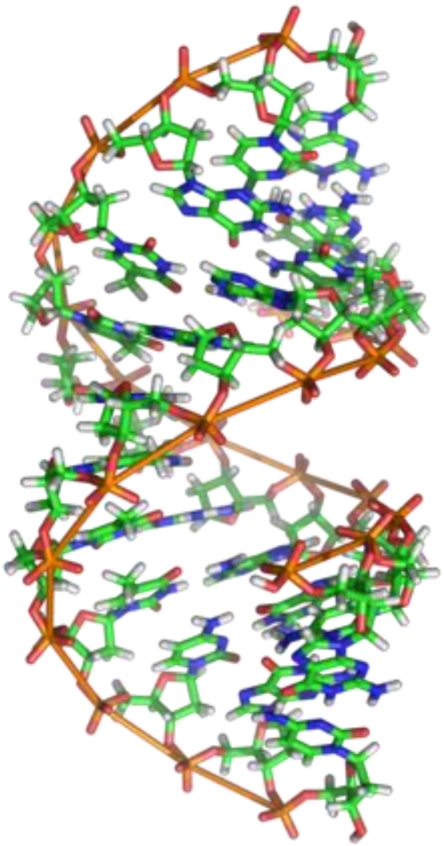


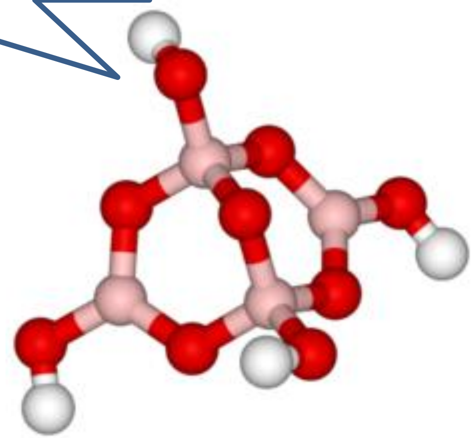


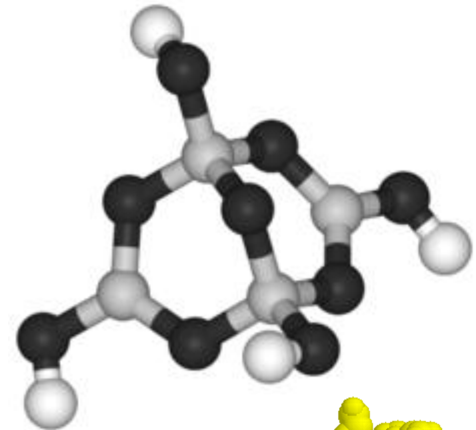
http://en.wikipedia.org/

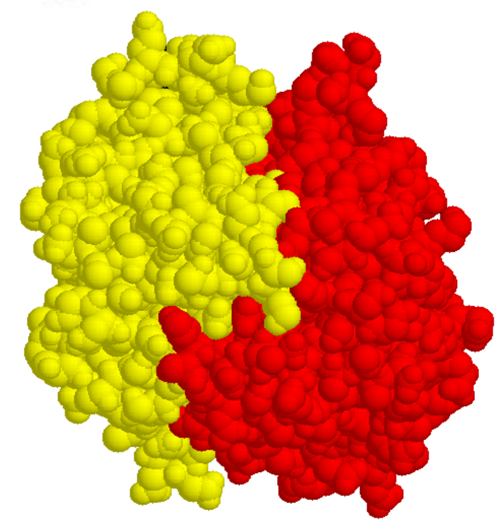
Bioinformatics

- Drug Discovery
- Gene Prioritization
- Protein Interaction Prediction
- ……





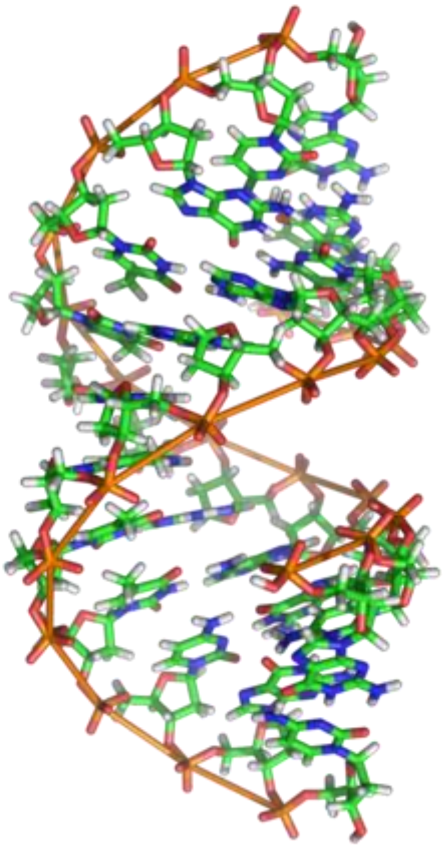


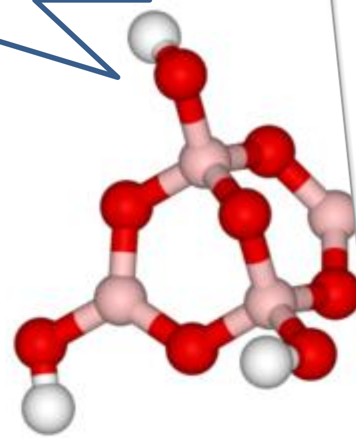


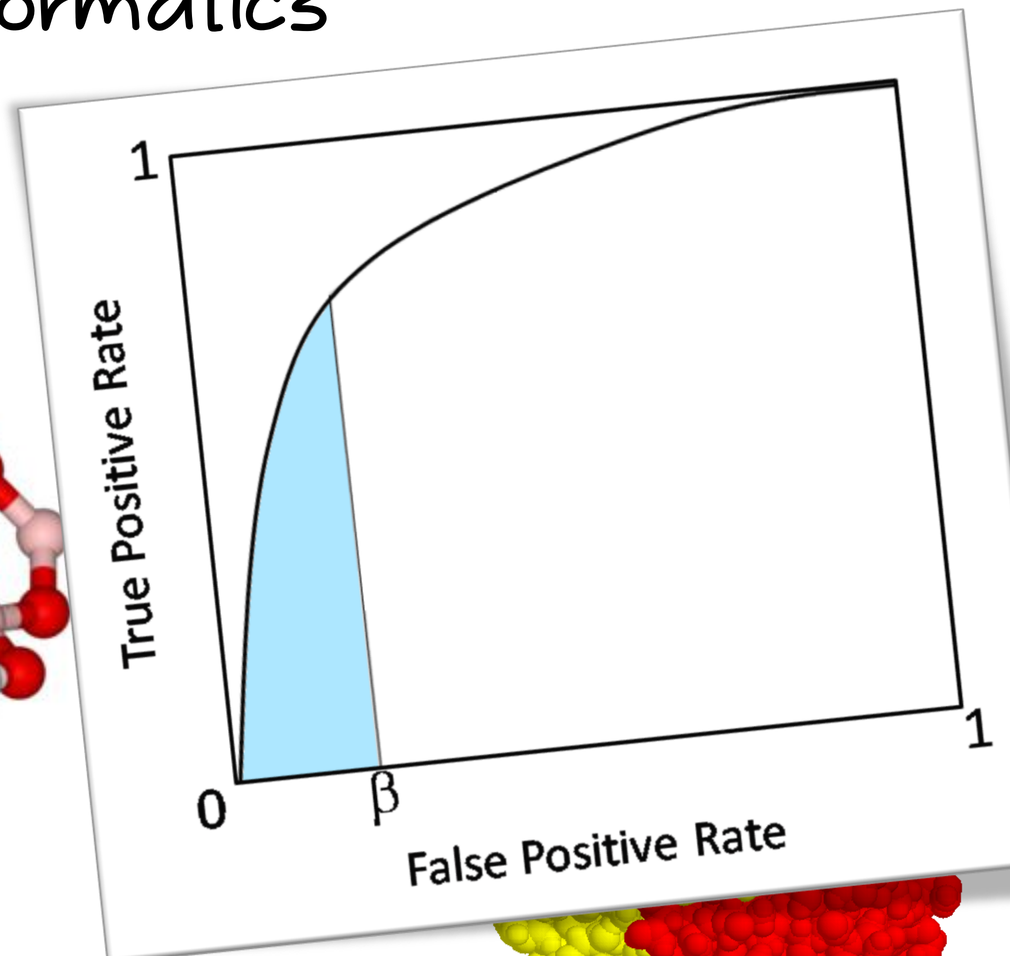
http://en.wikipedia.org/wiki
http://commons.wikimedia.org/
http://www.google.com/imghp

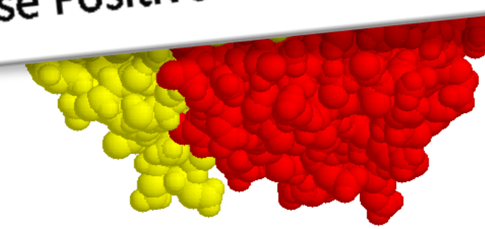
Bioinformatics

- Drug Discovery
- Gene Prioritization
- Protein Interaction Prediction
- ……



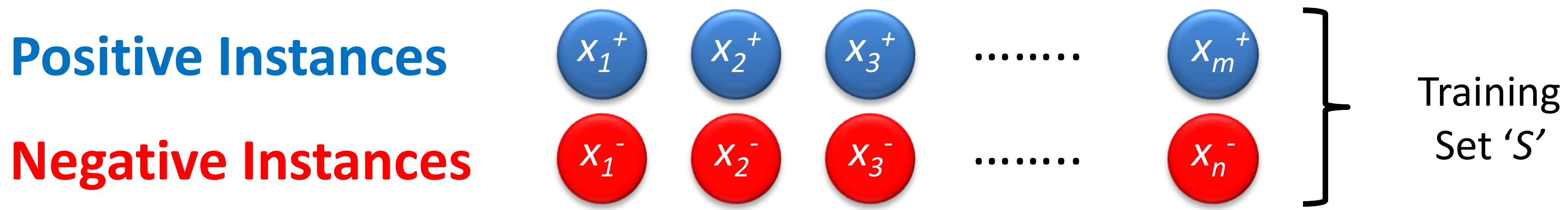






http://en.wikipedia.org/wiki
http://commons.wikimedia.org/
http://www.google.com/imghp

Partial AUC Optimization



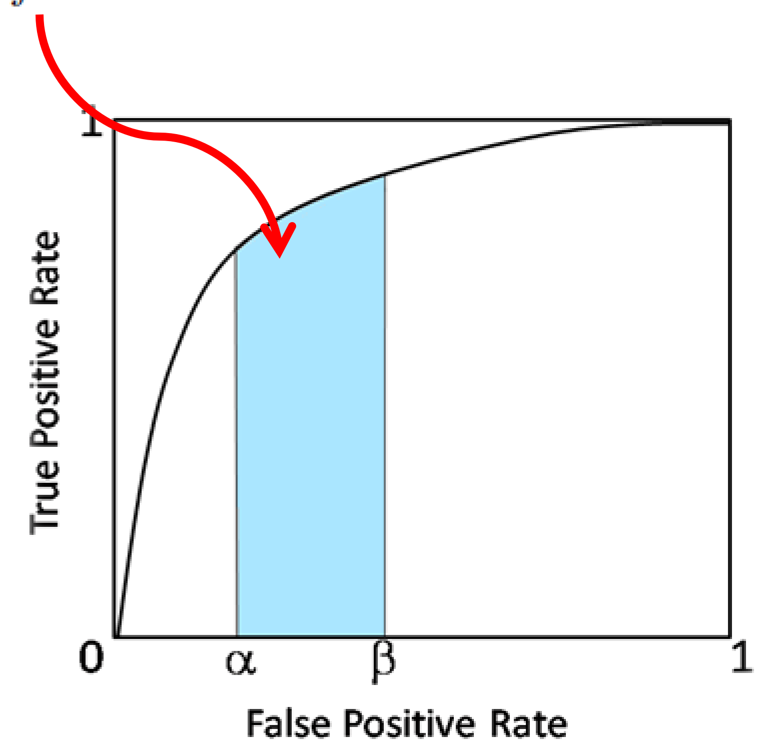
Partial AUC Optimization

Positive Instances x_1^+ x_2^+ x_3^+ …….. x_m^+

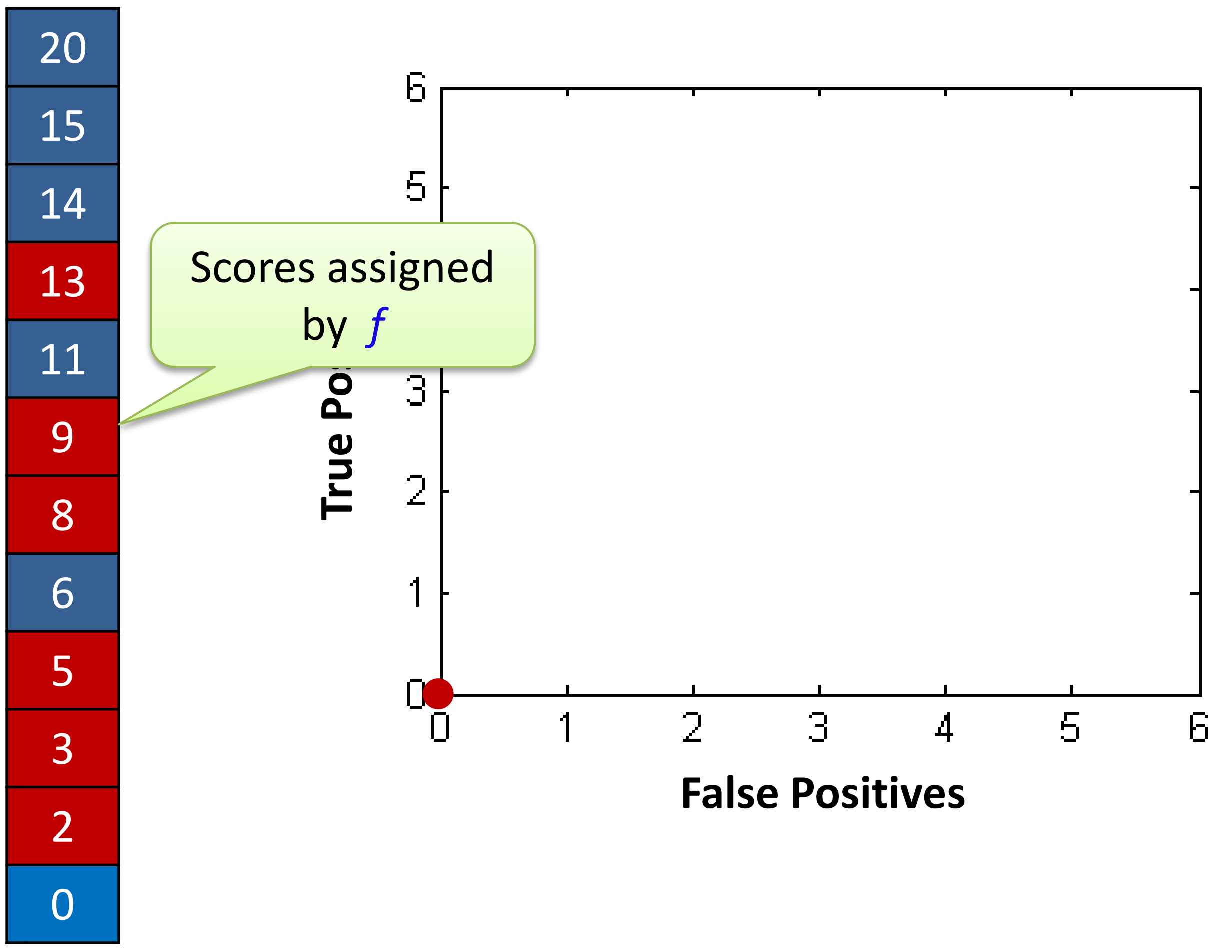
Negative Instances x_1^- x_2^- x_3^- …….. x_n^-

Training Set '*S*'

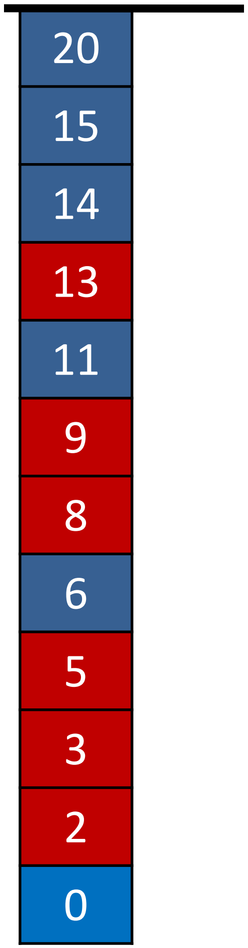
GOAL? Learn a scoring function $f : X \rightarrow \mathbb{R}$

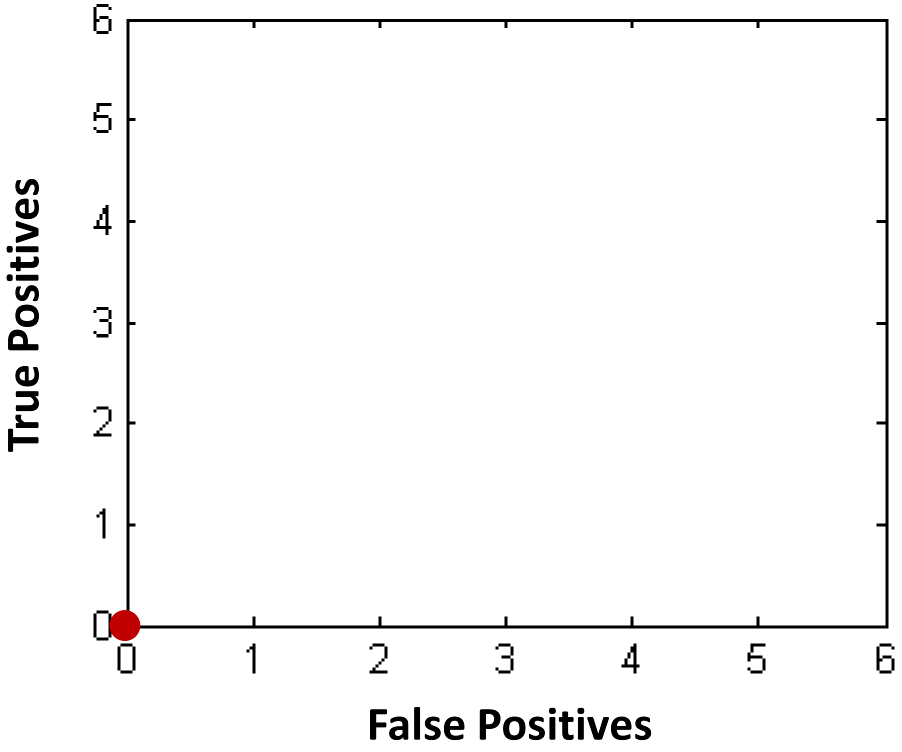


Receiver Operating Characteristic Curve Illustration

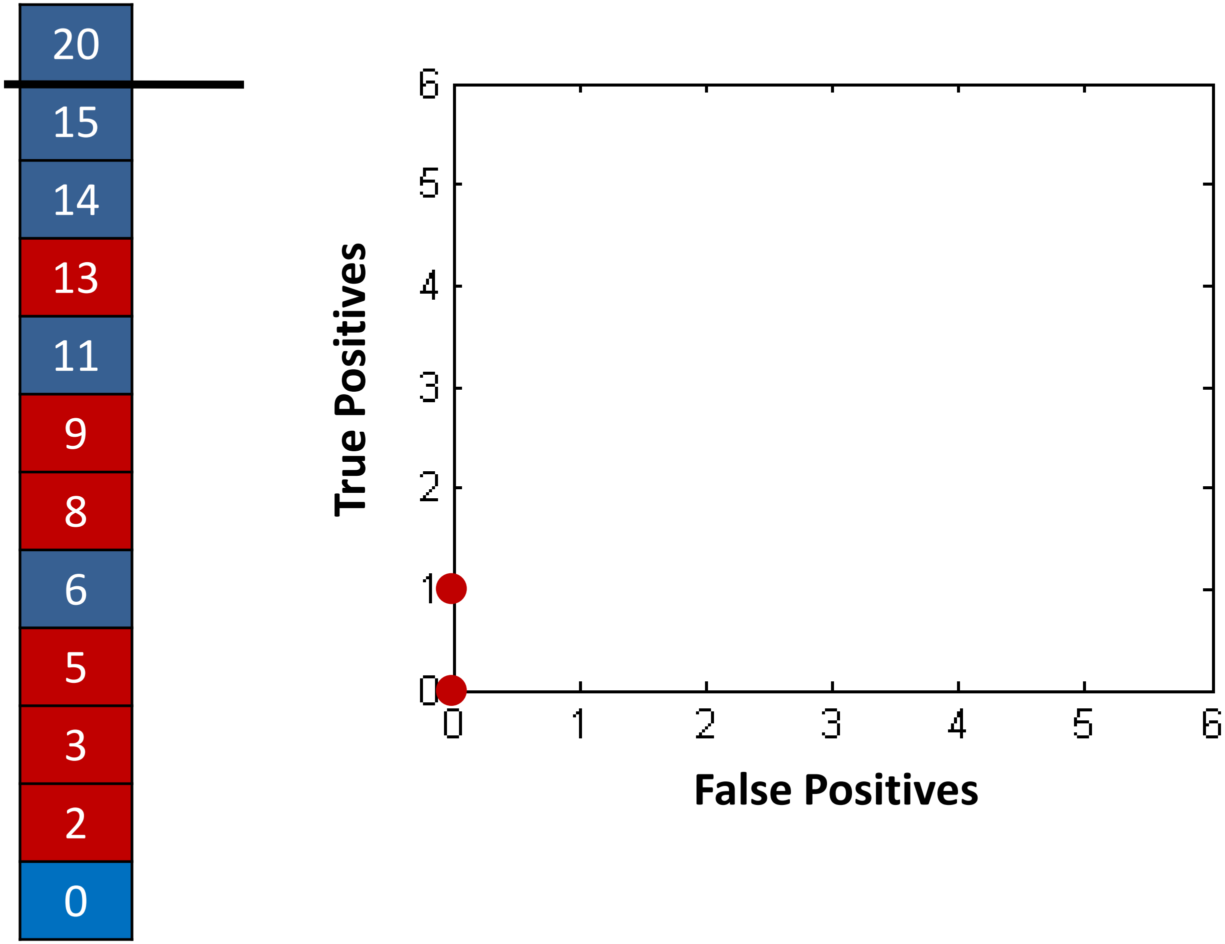


Receiver Operating Characteristic Curve Illustration

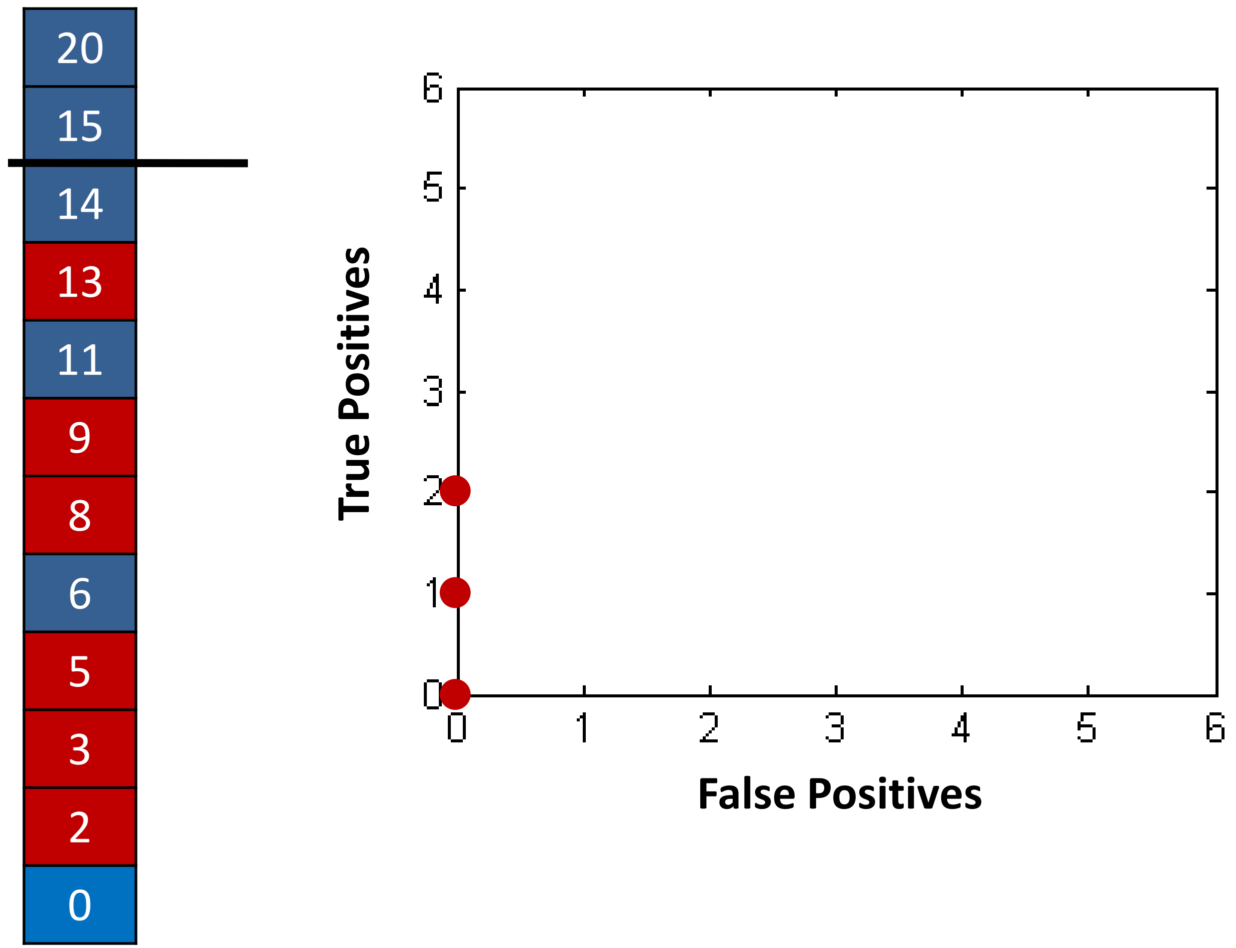




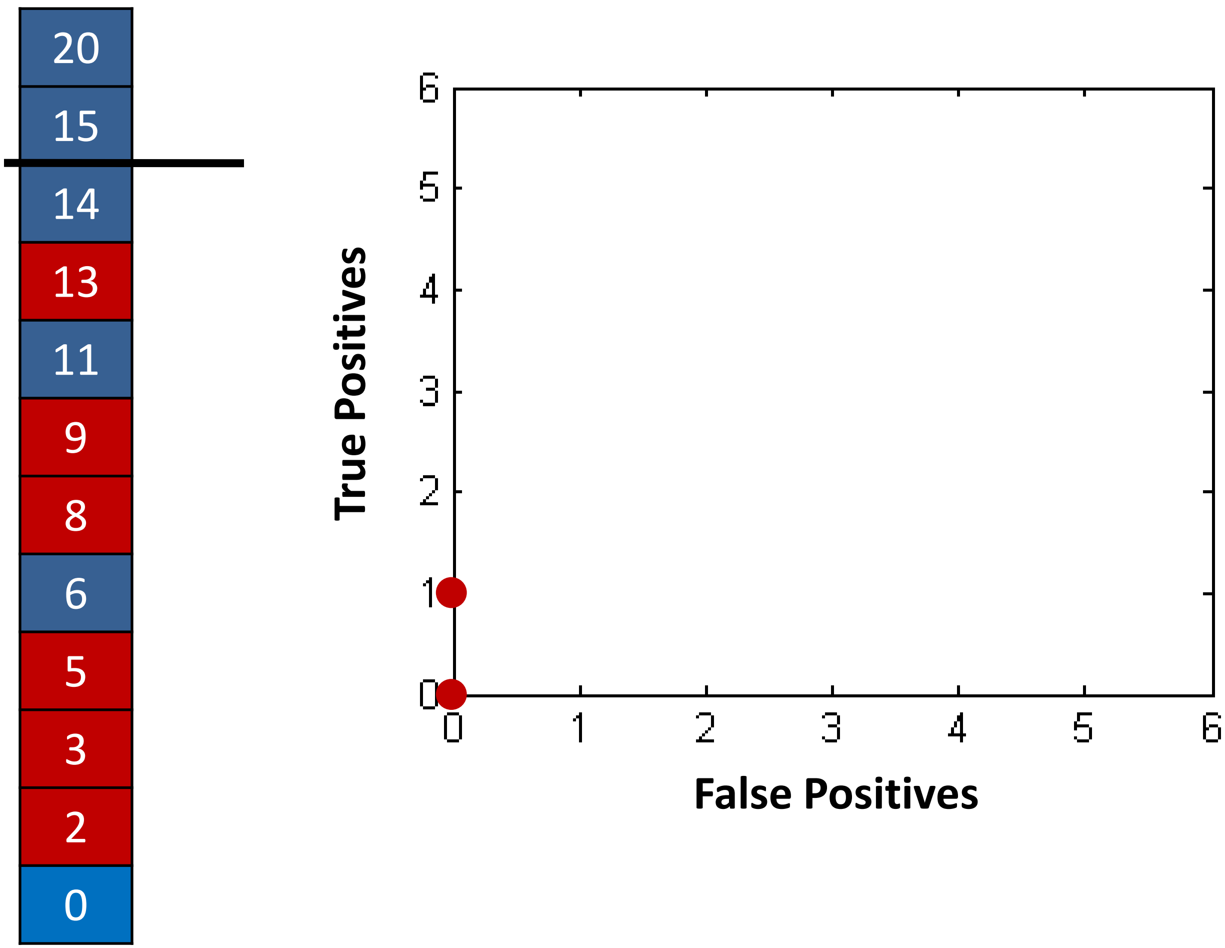
Receiver Operating Characteristic Curve Illustration



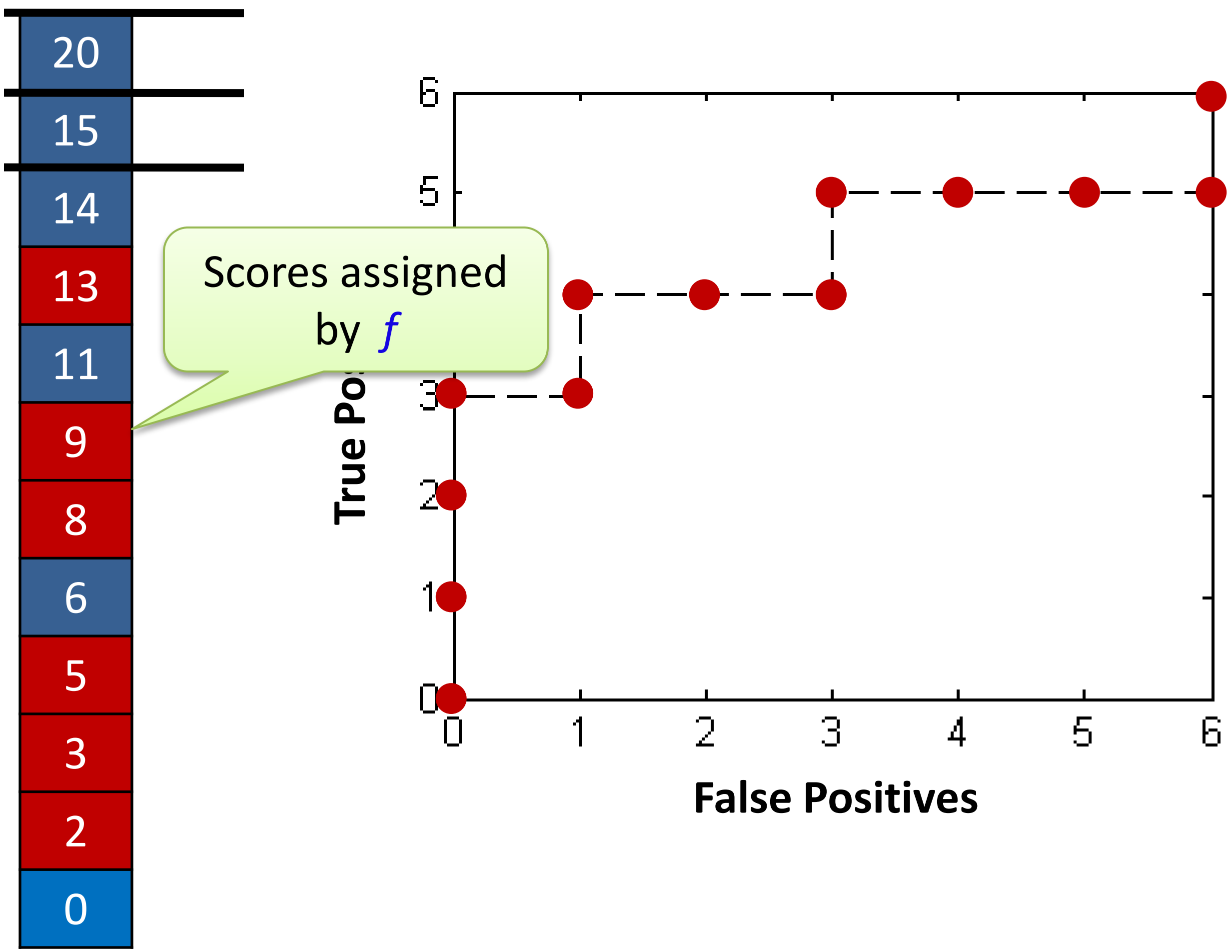
Receiver Operating Characteristic Curve Illustration



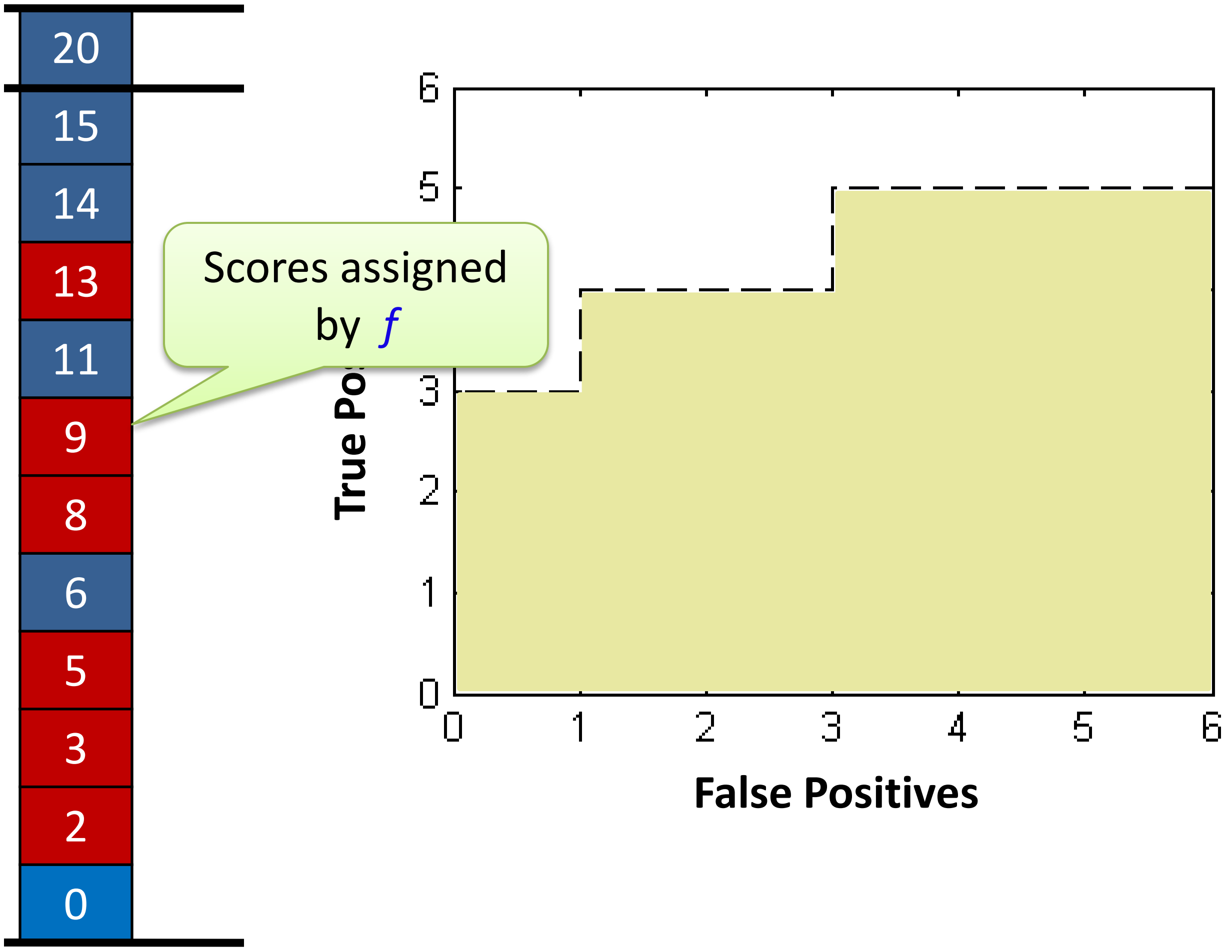
Receiver Operating Characteristic Curve Illustration



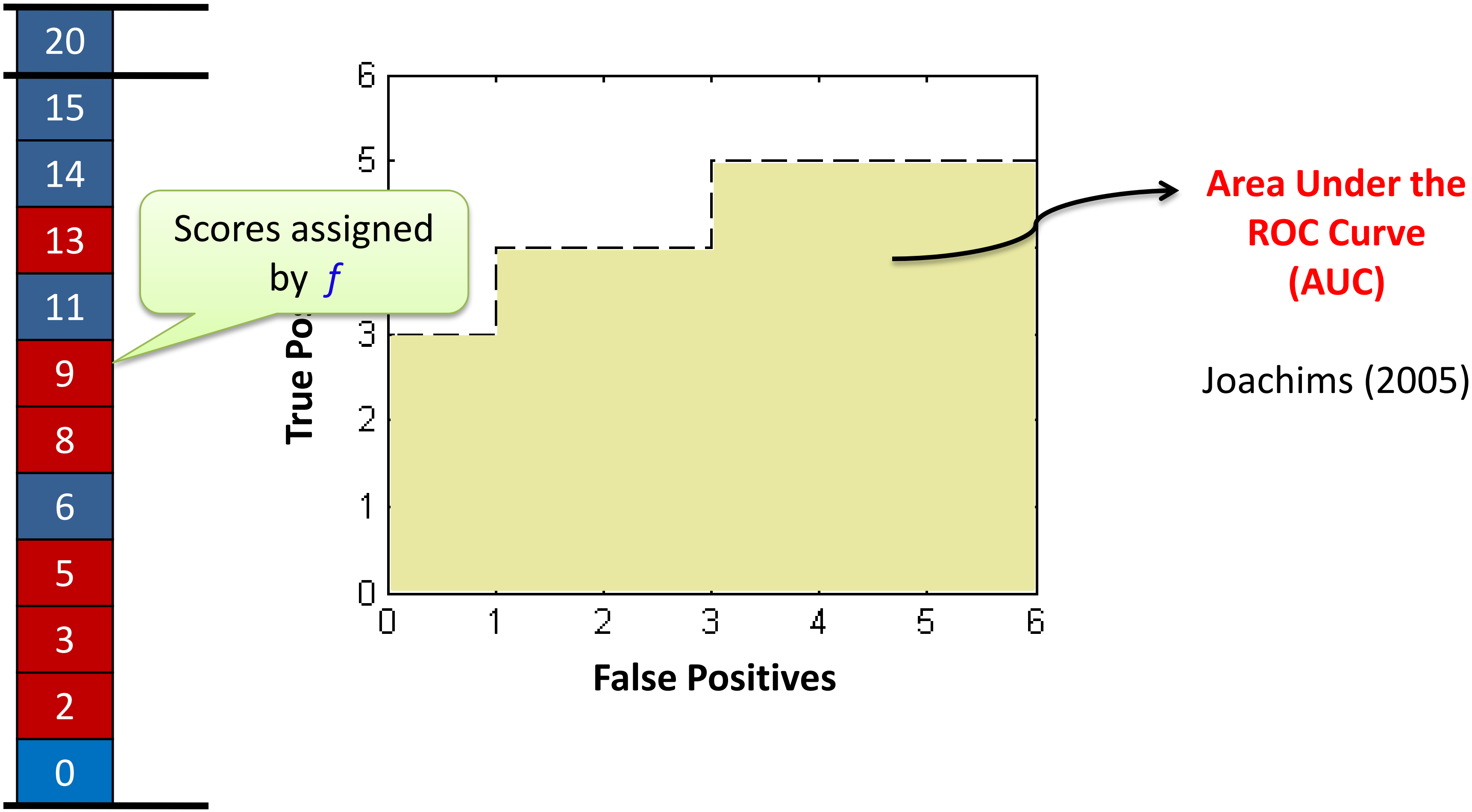
Receiver Operating Characteristic Curve Illustration



Receiver Operating Characteristic Curve Illustration

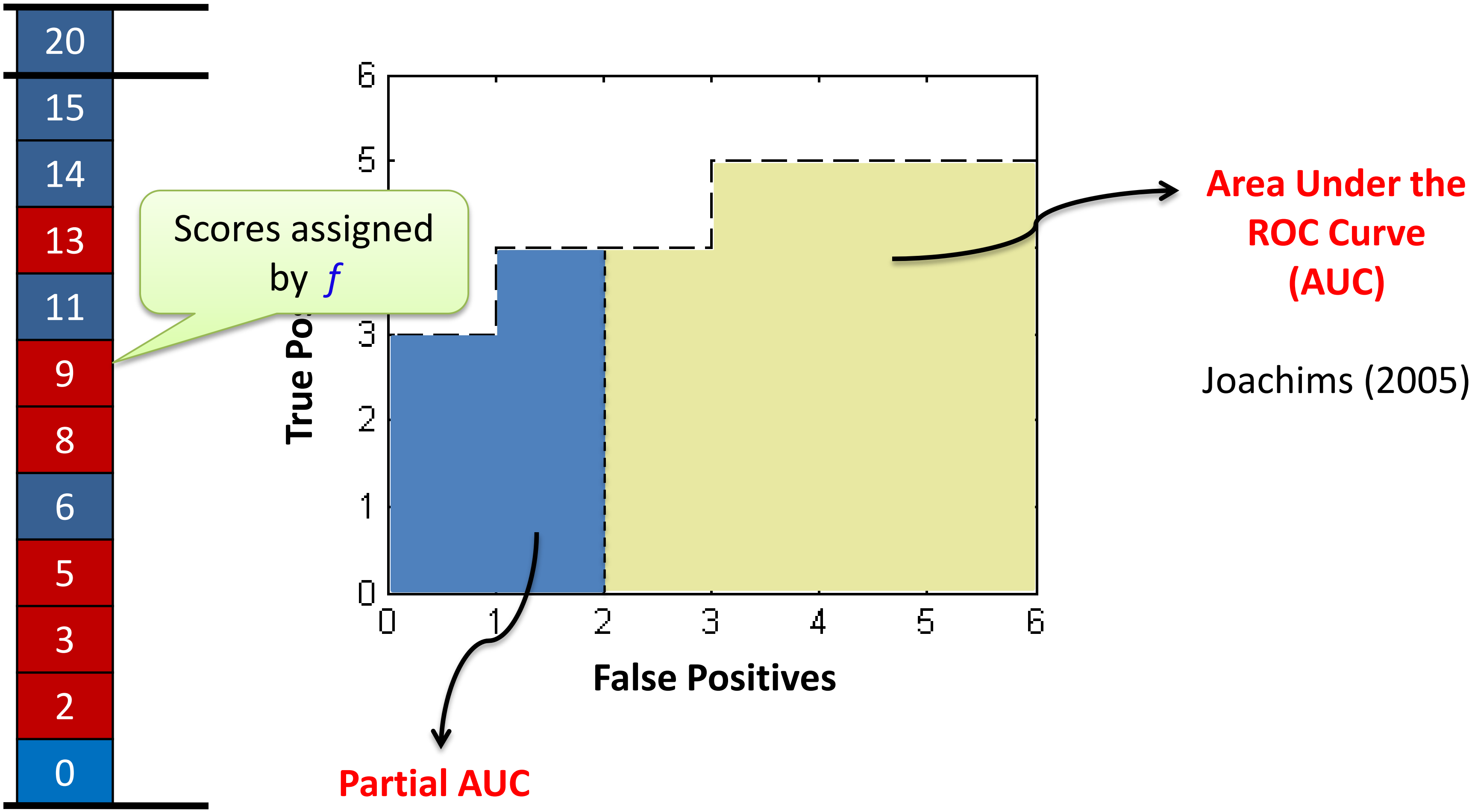


Receiver Operating Characteristic Curve Illustration



Joachims (2005)

Receiver Operating Characteristic Curve Illustration



Joachims (2005)

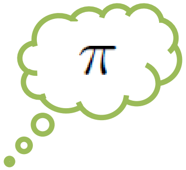
Partial AUC Optimization

Ordering of examples in *S*

n

m

0	0	0	0	0
1	1	0	0	0
1	1	0	0	1
1	1	0	0	1



Partial AUC Optimization

Ordering of examples in *S*

n

m

0	0	0	0	0
1	1	0	0	0
1	1	0	0	1
1	1	0	0	1

π

compared with

π^*

0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

IDEAL

Partial AUC Optimization

Ordering of examples in *S*

n

m

0	0	0	0	0
1	1	0	0	0
1	1	0	0	1
1	1	0	0	1

π

compared with

π^*

0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

IDEAL

$$\min_{w,\xi\geq 0} \frac{1}{2}||w||^2 + C\xi$$

s.t.

$$\forall \pi \in \Pi_{m,n}: \quad w^\top \left(\phi(S,\pi^*) - \phi(S,\pi)\right) \geq \Delta_{\mathrm{pAUC}(\alpha,\beta)}(\pi^*,\pi) - \xi$$

Partial AUC Optimization

Ordering of examples in *S*

π

$$\begin{array}{cc} & n \\ m & \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

compared with

π^*

IDEAL

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Upper Bound on (1 – pAUC)

$$\min_{w,\xi\geq 0} \frac{1}{2}||w||^2 + C\boxed{\xi}$$

s.t.

pAUC Loss

$$\forall \pi \in \Pi_{m,n} : \quad w^\top\left(\phi(S,\pi^*) - \phi(S,\pi)\right) \geq \Delta_{\mathrm{pAUC}(\alpha,\beta)}(\pi^*,\pi) - \xi$$

Cutting-plane Solver

$$\min_{w,\xi\geq 0} \frac{1}{2}||w||^2 + C\xi$$

$$\text{s.t. } \forall \pi \in \mathcal{C}:$$

$$\boxed{w^\top(\phi(S,\pi^*) - \phi(S,\pi)) \geq \Delta_{\text{pAUC}(\alpha,\beta)}(\pi^*,\pi) - \xi}$$

Repeat:

1. Solve OP for a subset of constraints.
2. Add the most violated constraint.

Cutting-plane Solver

$$\min_{w,\xi\geq 0} \frac{1}{2}||w||^2 + C\xi$$

s.t. $\forall \pi \in \mathcal{C}$:

$$\boxed{w^\top(\phi(S,\pi^*) - \phi(S,\pi)) \geq \Delta_{\mathrm{pAUC}(\alpha,\beta)}(\pi^*,\pi) - \xi}$$

Repeat:

1. Solve OP for a subset of constraints.
2. Add the most violated constraint.

Break down!

$$\operatorname*{argmax}_{\pi} \; \Delta_{\mathrm{pAUC}(\alpha,\beta)}(\pi^*,\pi) + w^\top(\phi(S,\pi^*) - \phi(S,\pi))$$

Cutting-plane Solver

$$\min_{w,\xi\geq 0} \frac{1}{2}||w||^2 + C\xi$$

s.t. $\forall \pi \in \mathcal{C}$:

$$\boxed{w^\top \left(\phi(S,\pi^*) - \phi(S,\pi)\right) \;\geq \Delta_{\text{pAUC}(\alpha,\beta)}(\pi^*,\pi) - \xi}$$

Repeat:

1. Solve OP for a subset of constraints.
2. Add the most violated constraint.

Break down!

$$\underset{\pi}{\text{argmax}} \;\; \Delta_{\text{pAUC}(\alpha,\beta)}(\pi^*,\pi) + \; w^\top \left(\phi(S,\pi^*) - \phi(S,\pi)\right)$$

Full AUC

0	1	0	1	0
1	1	0	0	0
1	1	0	0	1
1	1	0	0	1

Cutting-plane Solver

$$\min_{w,\xi\geq 0} \frac{1}{2}||w||^2 + C\xi$$

s.t. $\forall \pi \in \mathcal{C}$:

$$\boxed{w^\top\left(\phi(S,\pi^*) - \phi(S,\pi)\right) \;\geq \Delta_{\text{pAUC}(\alpha,\beta)}(\pi^*,\pi) - \xi}$$

Repeat:

1. Solve OP for a subset of constraints.
2. Add the most violated constraint.

Break down!

$$\underset{\pi}{\operatorname{argmax}} \;\; \Delta_{\text{pAUC}(\alpha,\beta)}(\pi^*,\pi) + \; w^\top\left(\phi(S,\pi^*) - \phi(S,\pi)\right)$$

Full AUC

0	1	0	1	0
1	1	0	0	0
1	1	0	0	1
1	1	0	0	1

Partial AUC

0	1	0	1	0
1	1	0	0	0
1	1	0	0	1
1	1	0	0	1

?

Trickier Optimization Problem

Full AUC

$$\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}1\left(f(x_i^+) > f(x_j^-)\right)$$

Trickier Optimization Problem

Full AUC

All Pairs

$$\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}1\left(f(x_i^+) > f(x_j^-)\right)$$

Trickier Optimization Problem

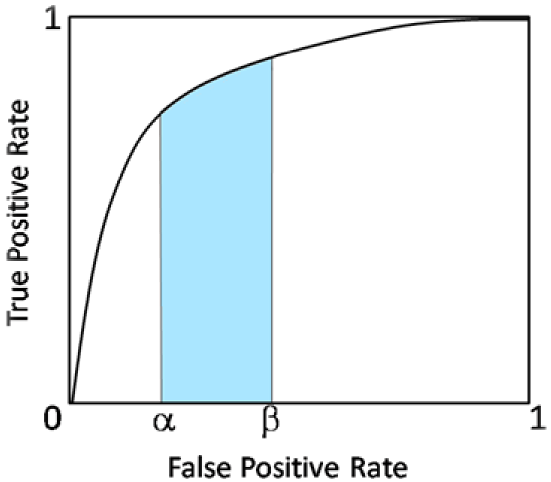
Full AUC

All Pairs

$$\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbb{1}\left(f(x_i^+) > f(x_j^-)\right)$$

Partial AUC

$$\frac{1}{mn(\beta-\alpha)}\sum_{i=1}^{m}\sum_{j=j_\alpha+1}^{j_\beta}\mathbb{1}\left(f(x_i^+) > f(x_{(j)}^-)\right)$$



Trickier Optimization Problem

Full AUC

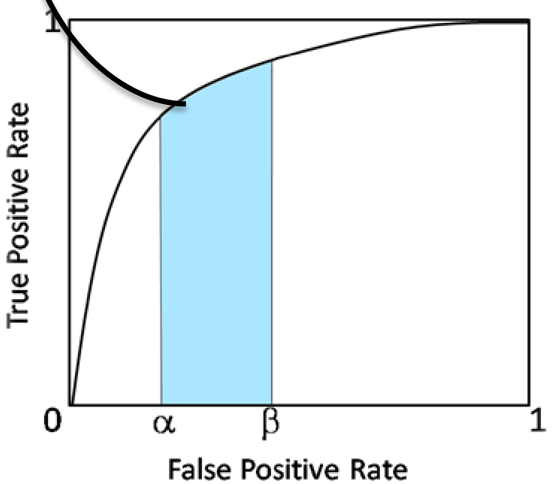
All Pairs

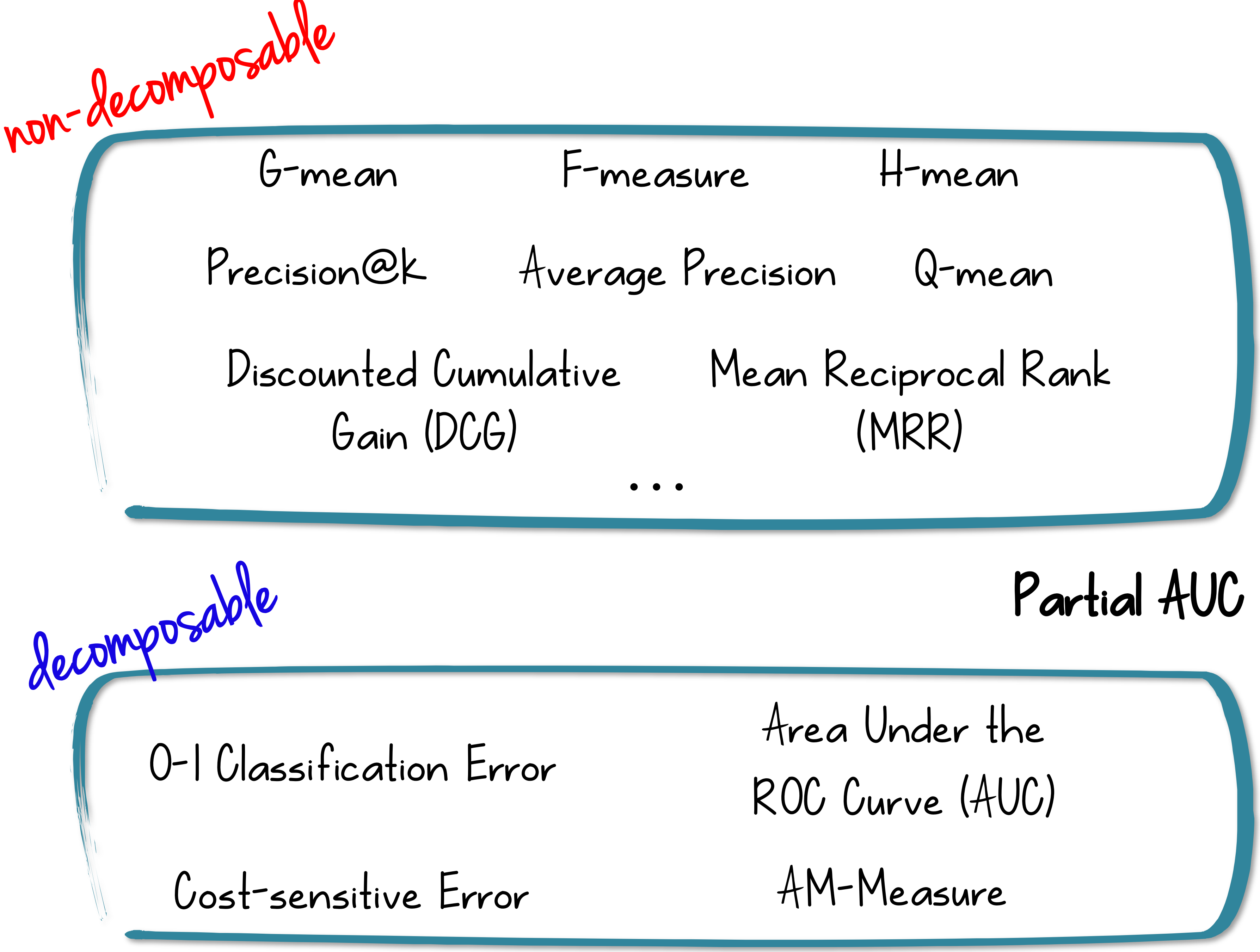
$$\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{1}\left(f(x_i^+) > f(x_j^-)\right)$$

Partial AUC

$$\frac{1}{mn(\beta-\alpha)}\sum_{i=1}^{m}\sum_{j=j_\alpha+1}^{j_\beta}\mathbf{1}\left(f(x_i^+) > f(x_{(j)}^-)\right)$$

Subset of negative instances in the FPR range [α, β] – **changes with ordering**



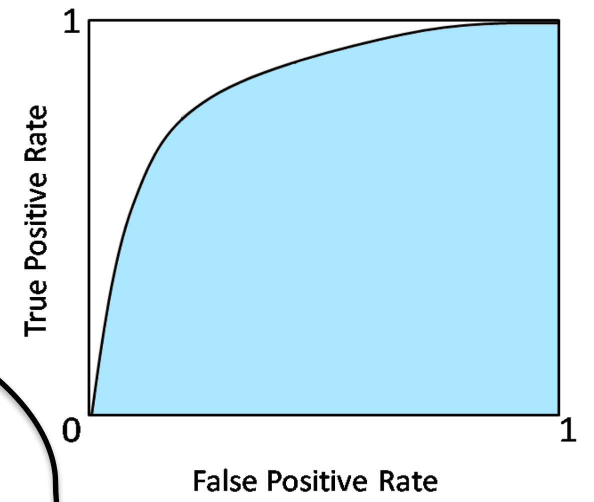

non-decomposable
G-mean
F-measure
H-mean
Precision@k
Average Precision
Q-mean
Discounted Cumulative Gain (DCG)
Mean Reciprocal Rank (MRR)
...
Partial AUC
decomposable
0-1 Classification Error
Area Under the ROC Curve (AUC)
Cost-sensitive Error
AM-Measure

Trickier Optimization Problem

Full AUC

All Pairs

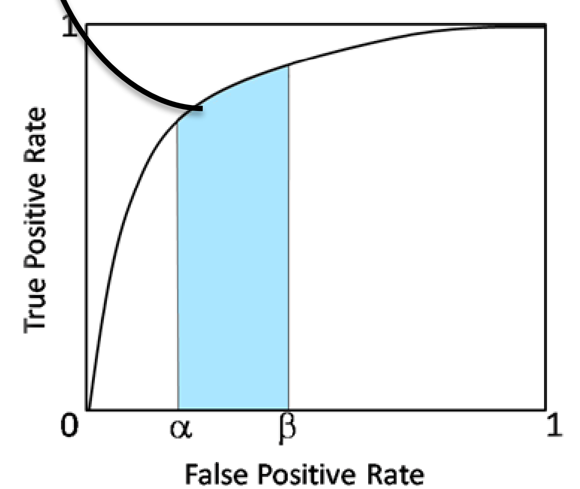
$$\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbb{1}\left(f(x_i^+) > f(x_j^-)\right)$$



Partial AUC

$$\frac{1}{mn(\beta-\alpha)}\sum_{i=1}^{m}\sum_{j=j_\alpha+1}^{j_\beta}\mathbb{1}\left(f(x_i^+) > f(x_{(j)}^-)\right)$$

Subset of negative instances in the FPR range [α, β] – **changes with ordering**



Trickier Optimization Problem

Full AUC

All Pairs

$$\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{1}(f(x_i^+) > f(x_j^-))$$

Partial AUC

$$\frac{1}{mn(\beta-\alpha)}\sum_{i=1}^{m}\sum_{j=j_\alpha+1}^{j_\beta}\mathbf{1}(f(x_i^+) > f(x_{(j)}^-))$$

Subset of negative instances in the FPR range [α, β] – **changes with ordering**

Partial AUC

0	1	0	1	0
1	1	0	0	0
1	1	0	0	1
1	1	0	0	1

Optimize rows independently

0	1	0	1	0
1	1	0	0	0
1	1	0	0	1
1	1	0	0	1

Trickier Optimi

Can be implemented in **O((m+n) log (m+n))** time complexity

Inputs: $S = (S_+, S_-)$, α, β, w

For $i = 1, \ldots, m$ **do**

Optimize over $r_i \in \{0, \ldots, j_\alpha - 1\}$:

$$\pi^{(1)}_{i,(j)_w} = \begin{cases} \mathbf{1}(w^\top x^{\pm}_{i,(j)_w} \le 0)\,, & j \in \{1, \ldots, j_\alpha - 1\} \\ 0\,, & j \in \{j_\alpha, \ldots, n\} \end{cases}$$

Optimize over $r_i \in \{j_\alpha\}$:

$$\pi^{(2)}_{i,(j)_w} = \begin{cases} 1\,, & j \in \{1, \ldots, j_\alpha\} \\ 0\,, & j \in \{j_\alpha + 1, \ldots, n\} \end{cases}$$

Optimize over $r_i \in \{j_\alpha + 1, \ldots, n\}$:

$$\pi^{(3)}_{i,(j)_w} = \begin{cases} 1\,, & j \in \{1, \ldots, j_\alpha + 1\} \\ \mathbf{1}(w^\top x^{\pm}_{i,(j)_w} \le 1)\,, & j \in \{j_\alpha + 2, \ldots, j_\beta\} \\ \mathbf{1}(w^\top x^{\pm}_{i,(j)_w} \le n\beta - j_\beta)\,, & j = j_\beta + 1 \\ \mathbf{1}(w^\top x^{\pm}_{i,(j)_w} \le 0)\,, & j \in \{j_\beta + 2, \ldots, n\} \end{cases}$$

$$\bar{k} = \operatorname*{argmax}_{k \in \{1,2,3\}} \left\{ \begin{array}{l} \text{term inside sum over } i \text{ in} \\ \text{Eq. (4) evaluated at } \pi_i^{(k)} \end{array} \right\}$$

$\bar{\pi}_i = \pi_i^{(\bar{k})}$

End For

Output: $\bar{\pi}$

Partial AUC

0	1	0	1	0
1	1	0	0	0
1	1	0	0	1
1	1	0	0	1

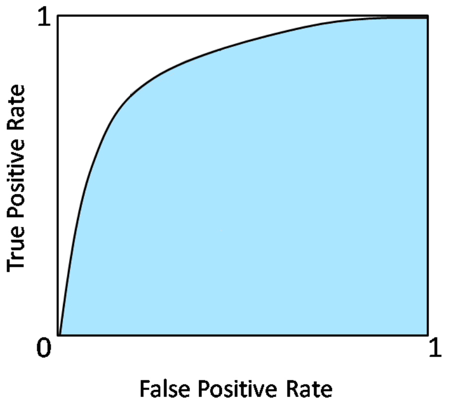
Optimize rows independently

0	1	0	1	0
1	1	0	0	0
1	1	0	0	1
1	1	0	0	1

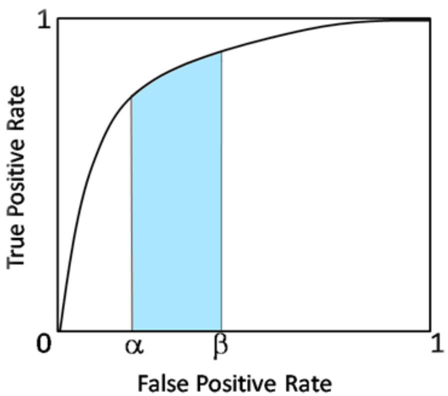
H. Narasimhan and S. Agarwal. *A Structural SVM Based Approach for Optimizing Partial AUC*. ICML, 2013.

Experimental Results

- Baseline Methods:
  - Full AUC Optimization (Joachims, 2005)

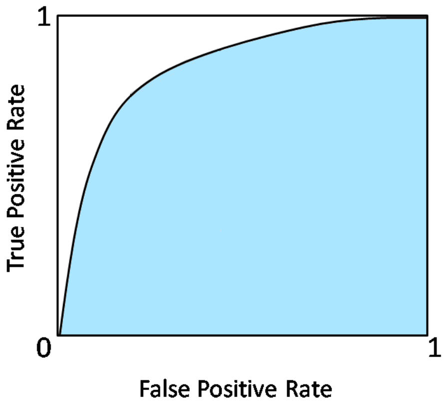


Vs

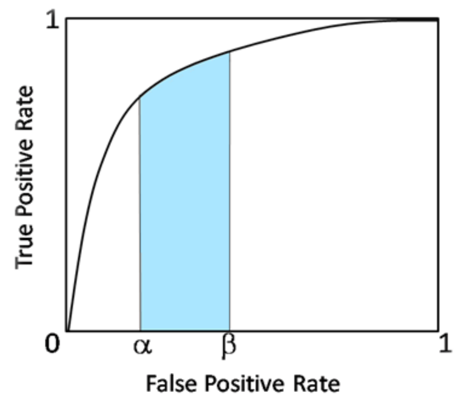


Experimental Results

- Baseline Methods:
  - Full AUC Optimization (Joachims, 2005)



Vs

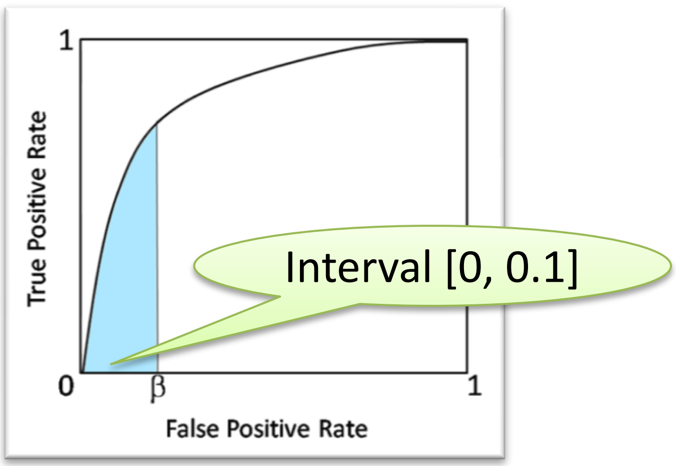


  - Asymmetric SVM (Wu et al., 2008)
  - Boosting Style Method (Komori & Eguchi, 2010)
  - Greedy Heuristic Method (Ricamato & Tortorella, 2011)

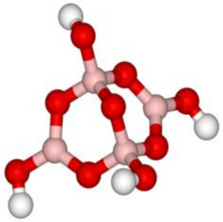
Experimental Results

Drug Discovery

50 active compounds / 2092 inactive compounds



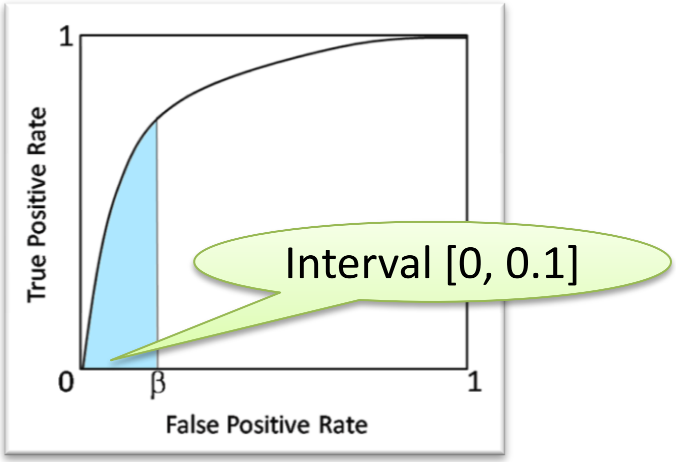
	Partial AUC in [0, 0.1]
SVMpAUC	**65.25**
SVM-AUC	62.64
ASVM	63.80
pAUCBoost	43.89
Greedy Heuristic	8.33



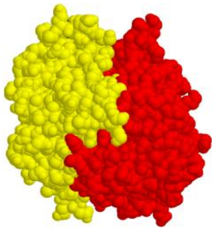
Experimental Results

Protein-Protein Interaction Prediction

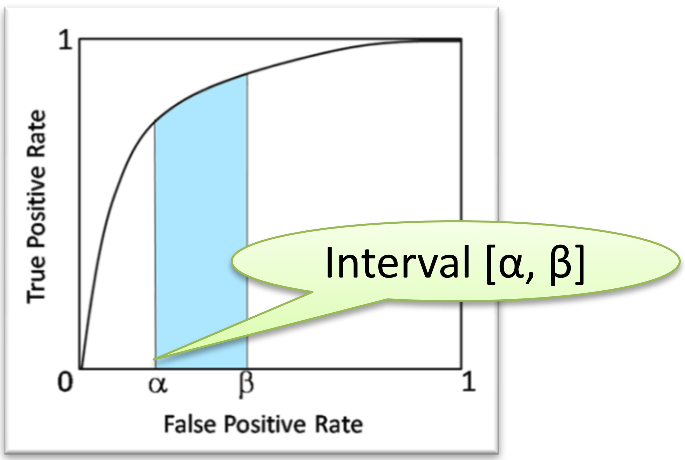
~3x10^3 interacting pairs / ~2x10^5 non-interacting pairs



	Partial AUC in [0, 0.1]
SVMpAUC	**51.79**
SVM-AUC	39.72
ASVM	44.51
pAUCBoost	48.65
Greedy Heuristic	47.33



Experimental Results



KDD Cup 2008

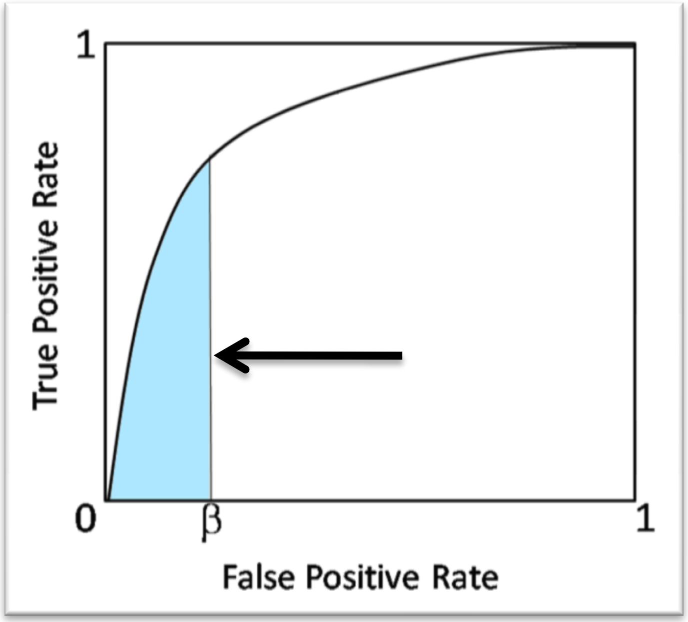
Breast Cancer Detection

~600 malignant ROIs / ~10^5 benign ROIs

	Partial AUC in [0.2s, 0.3s]
SVMpAUC	**51.44**
SVM-AUC	50.50
pAUCBoost	48.06
Greedy Heuristic	46.99

Experimental Results

Run Time Analysis



Cutting-plane Method

Repeat:

1. Solve OP for a subset of constraints.
2. Add the most violated constraint.

Experimental Results

Run Time Analysis

Cutting-plane Method

Repeat:

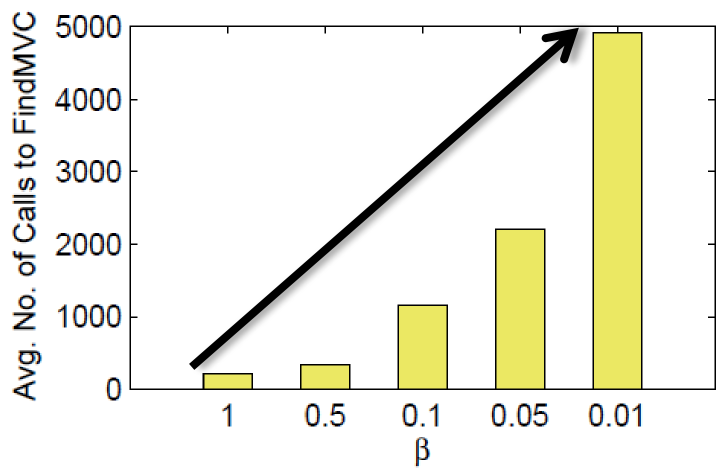
1. Solve OP for a subset of constraints.
2. Add the most violated constraint.

Time taken per iteration
Total number of iterations

Experimental Results

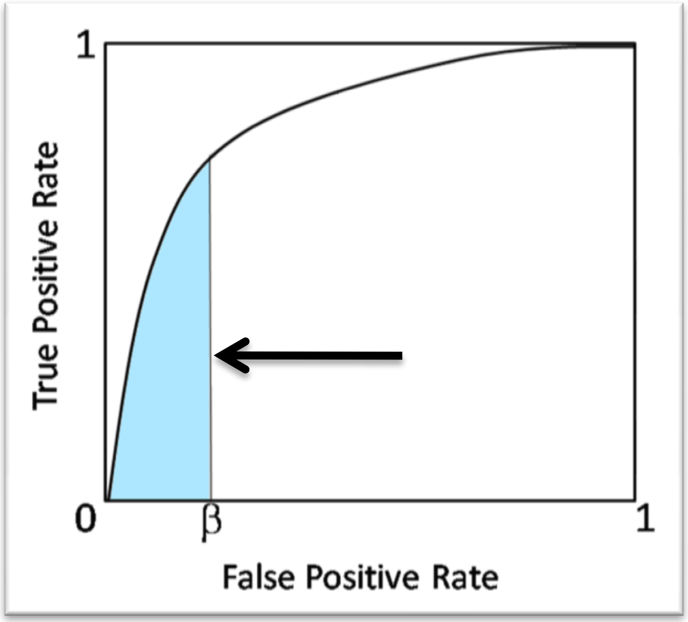
Run Time Analysis

Total number of iterations

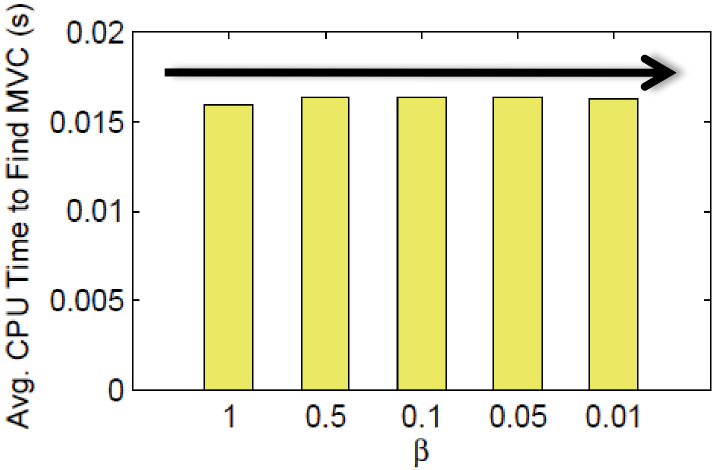


Experimental Results

Run Time Analysis



Time taken per iteration



Improved Formulation

- Characterize the Structural SVM Objective

Narasimhan, H. and Agarwal, S. *“SVM_pAUC^tight: A new support vector method for optimizing partial AUC based on a tight convex upper bound”*, KDD 2013.

Improved Formulation

- Characterize the Structural SVM Objective
- Better Formulation: Tighter Approximation

Narasimhan, H. and Agarwal, S. *"SVM_pAUC^tight: A new support vector method for optimizing partial AUC based on a tight convex upper bound"*, KDD 2013.

Improved Formulation

- Characterize the Structural SVM Objective
- Better Formulation: Tighter Approximation
  - Improved Accuracy
  - Better Run-time Guarantee

Narasimhan, H. and Agarwal, S. *"SVM_pAUC^tight: A new support vector method for optimizing partial AUC based on a tight convex upper bound"*, KDD 2013.

Algorithms

Two Approaches

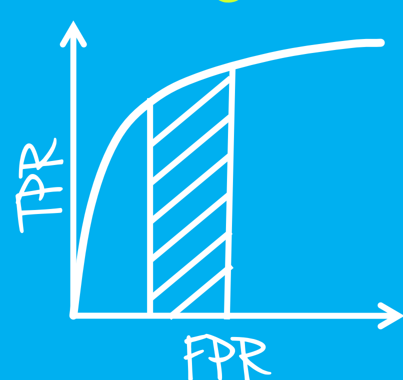
Plug-in

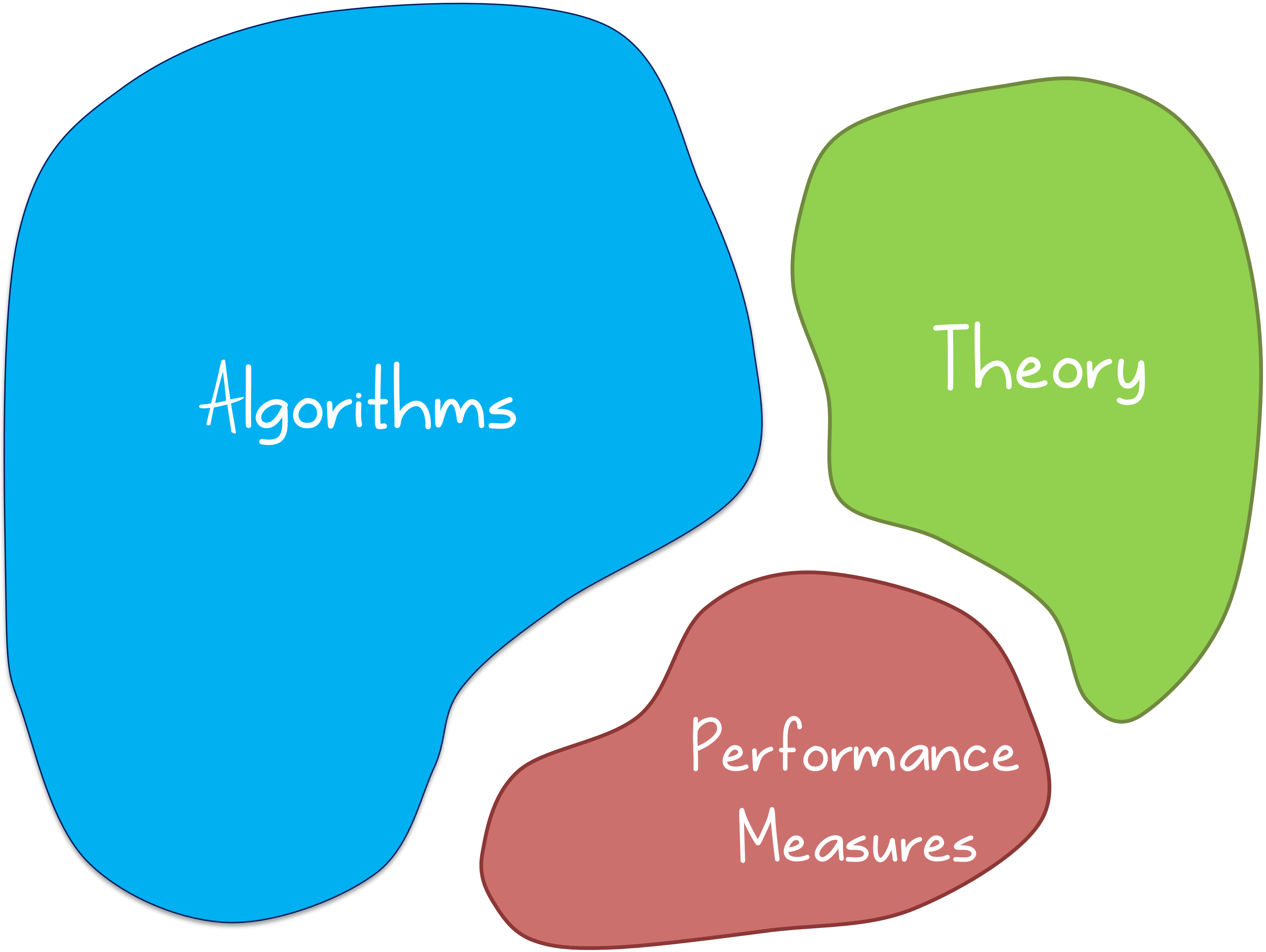
Risk Minimization

min {convex upper bound on loss}

Structural SVM

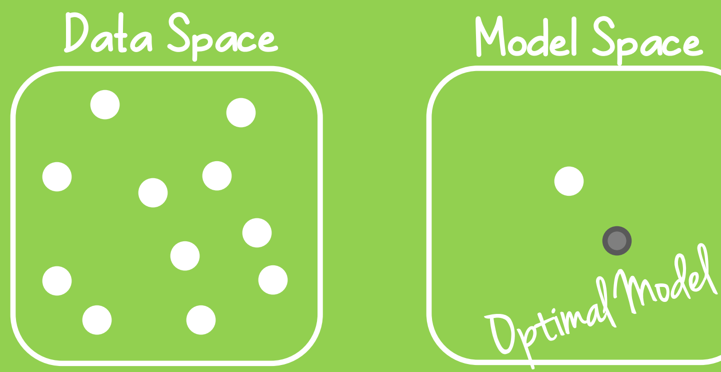
Case Study - Partial AUC




Algorithms
Theory
Performance
Measures

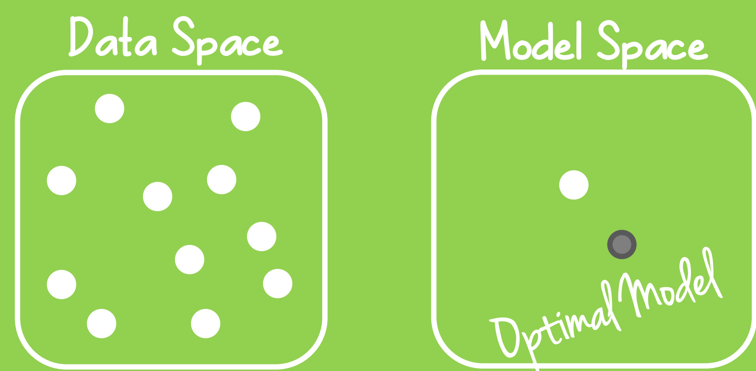
Theory

Statistical Consistency

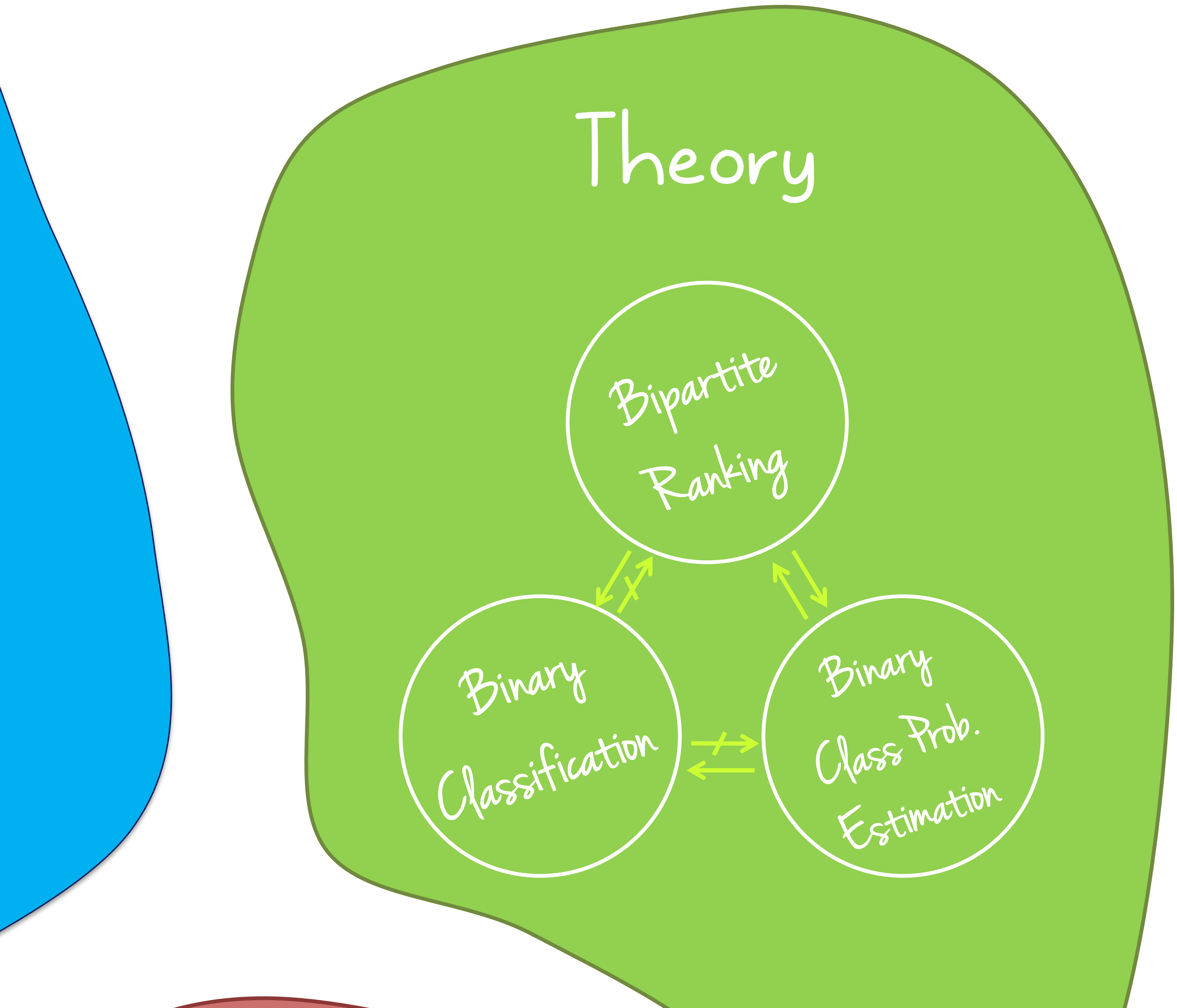


Theory

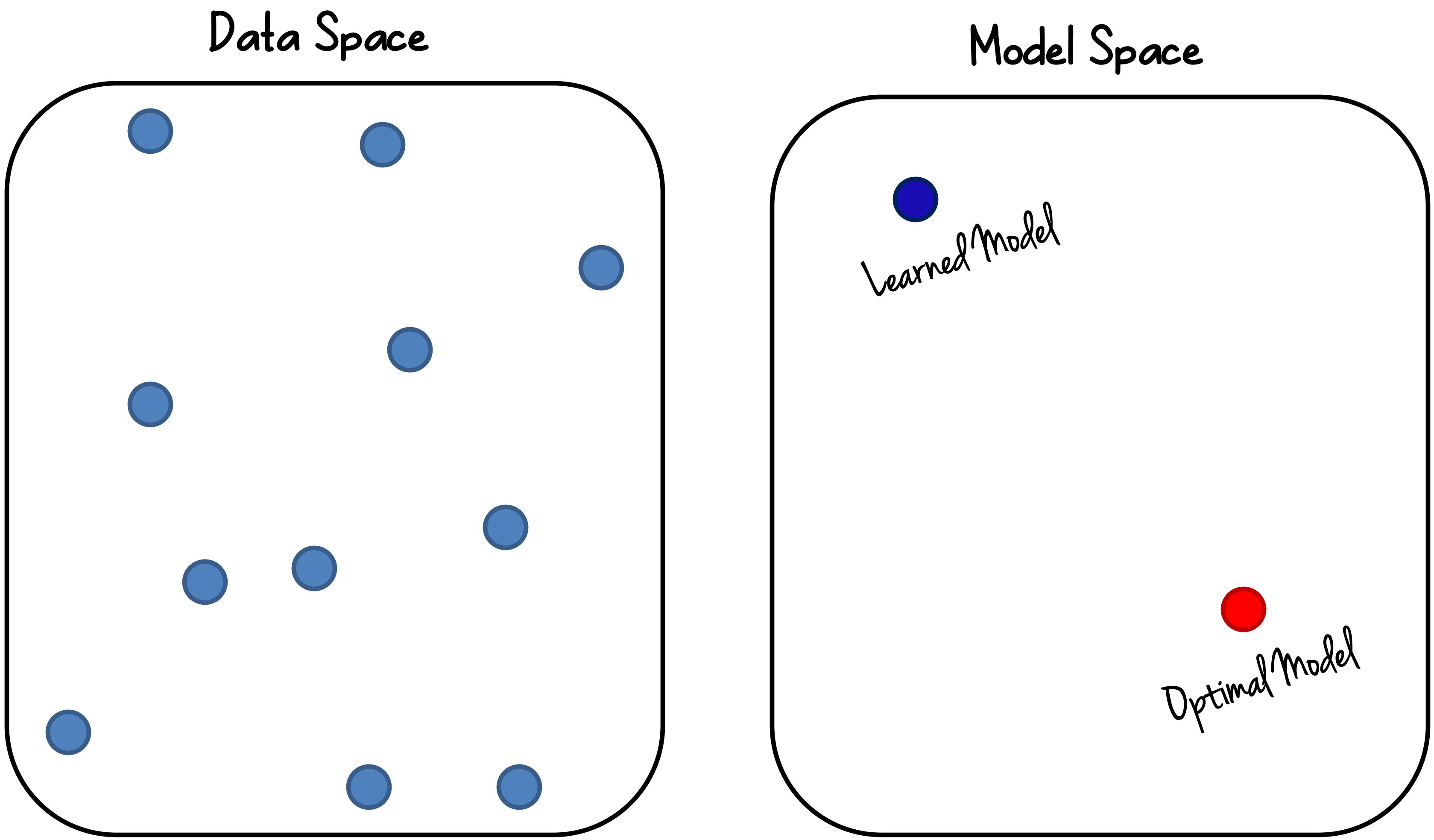
Statistical Consistency



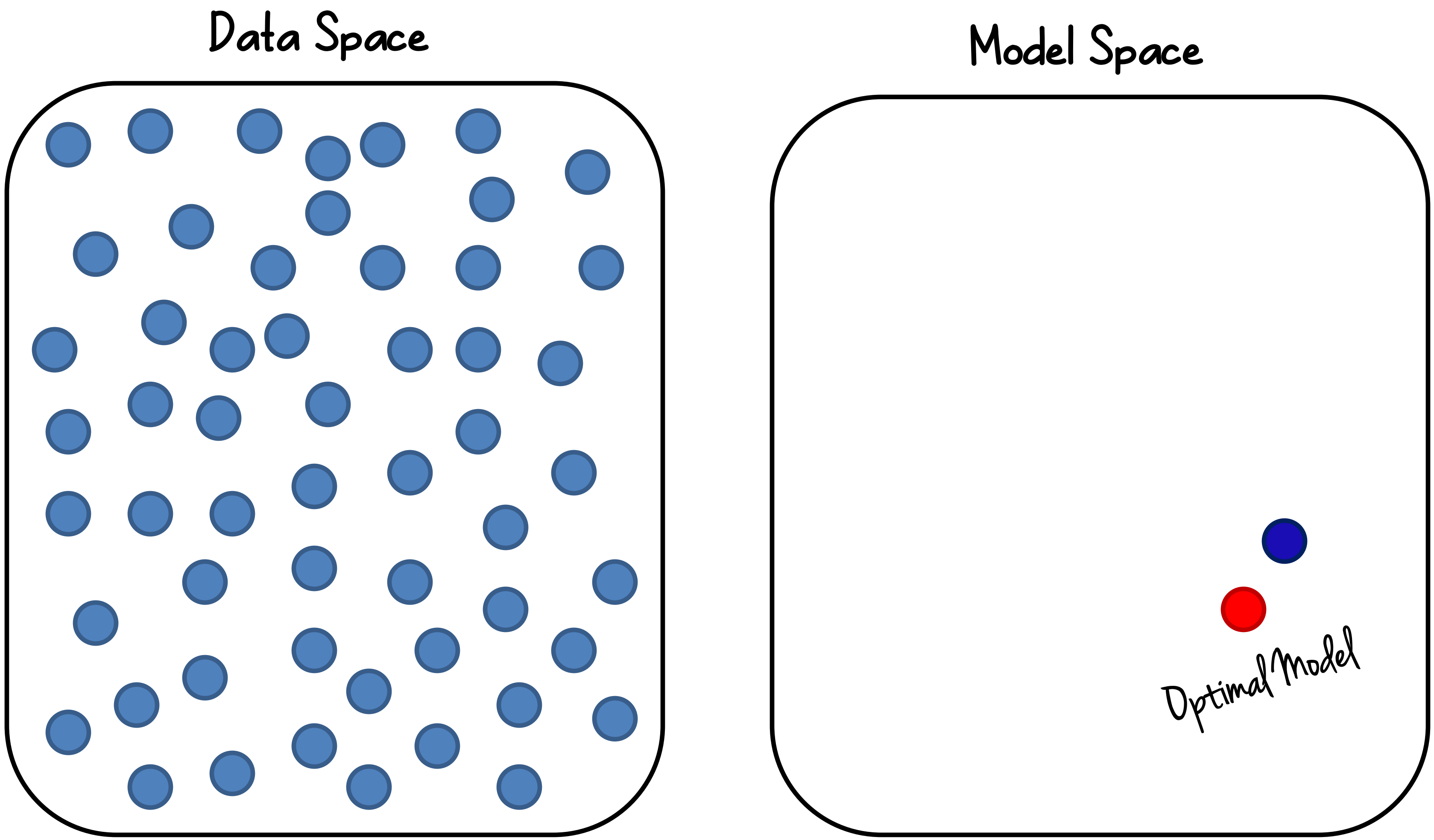
Case Study - Class Imbalance


Theory
Bipartite Ranking
Binary Classification
Binary Class Prob. Estimation

Statistical Consistency



Statistical Consistency



Statistical Consistency

$$\mathrm{er}_D^{0\text{-}1}[h] \;=\; \mathbf{E}_{(x,y)\sim D}\big[\mathbf{1}(h(x) \neq y)\big]$$

Statistical Consistency

$$\mathrm{er}_D^{0\text{-}1}[h] \;=\; \mathbf{E}_{(x,y)\sim D}\big[\mathbf{1}(h(x) \neq y)\big]$$

$$\mathrm{er}_D^{0\text{-}1,*} \;=\; \inf_{h:X\to\{\pm 1\}} \mathrm{er}_D^{0\text{-}1}[h]$$

Statistical Consistency

$$\mathrm{er}_D^{0\text{-}1}[h] \;=\; \mathbf{E}_{(x,y)\sim D}\big[\mathbf{1}(h(x) \neq y)\big]$$

$$\mathrm{er}_D^{0\text{-}1,*} \;=\; \inf_{h:X\to\{\pm 1\}} \mathrm{er}_D^{0\text{-}1}[h]$$

$$\boxed{\mathrm{regret}_D^{0\text{-}1}[h] \;=\; \mathrm{er}_D^{0\text{-}1}[h] - \mathrm{er}_D^{0\text{-}1,*}}$$

Statistical Consistency

$$S = ((x_1, y_1), \ldots, (x_n, y_n)) \in (X \times \{\pm 1\})^n$$

drawn iid from D

0-1 Consistency

$$\mathrm{regret}_D^{0\text{-}1}[h_S] \xrightarrow{P} 0$$

Statistical Consistency

$$S = ((x_1, y_1), \ldots, (x_n, y_n)) \in (X \times \{\pm 1\})^n$$

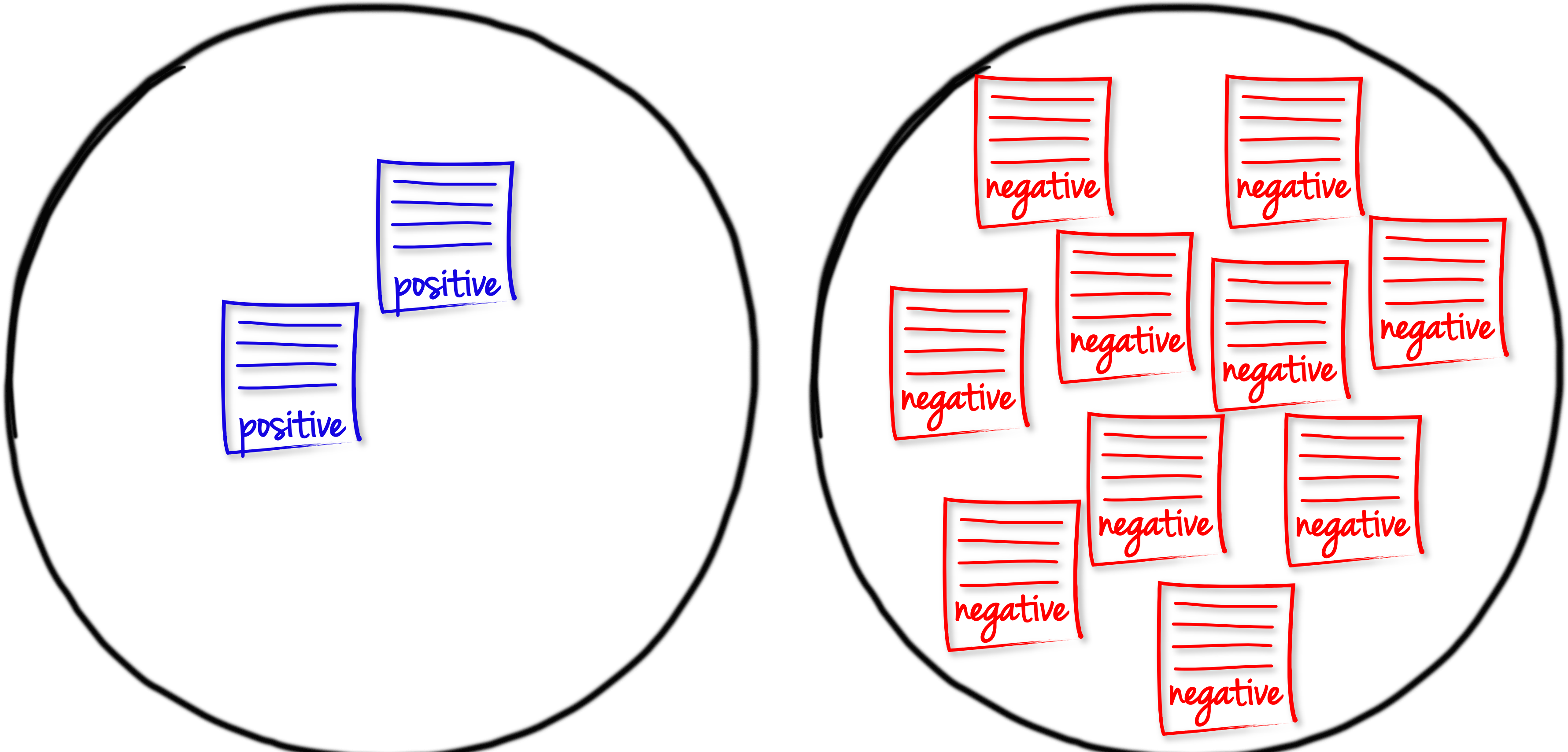
drawn iid from D

M-Consistency

$$\mathrm{regret}_D^{\mathrm{M}}[h_S] \xrightarrow{P} 0$$

Class Imbalance

$p = \mathbf{P}(y = 1)$ departs significantly from 0.5



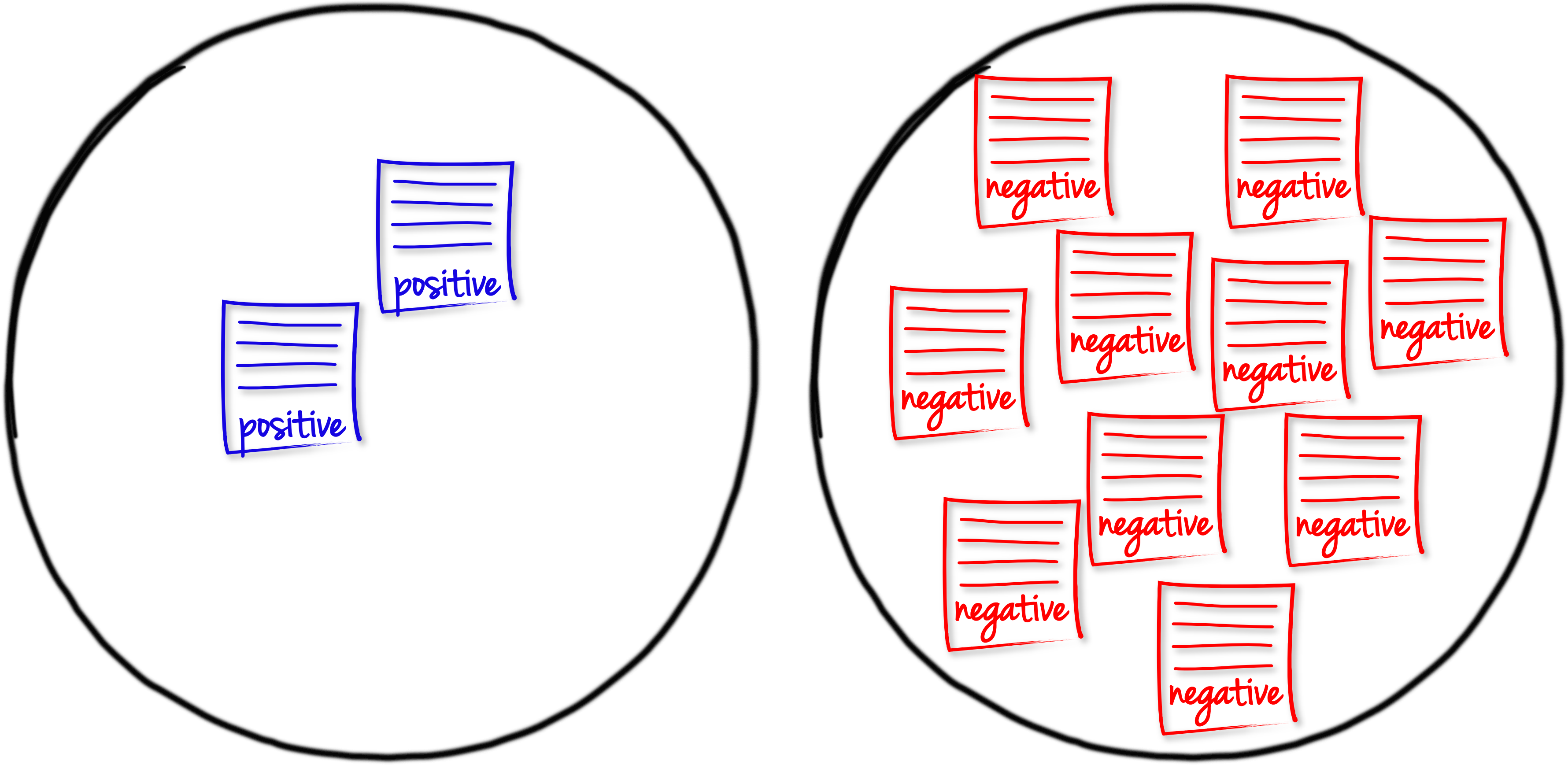
Standard Classification Error Ill-suited!

Class Imbalance

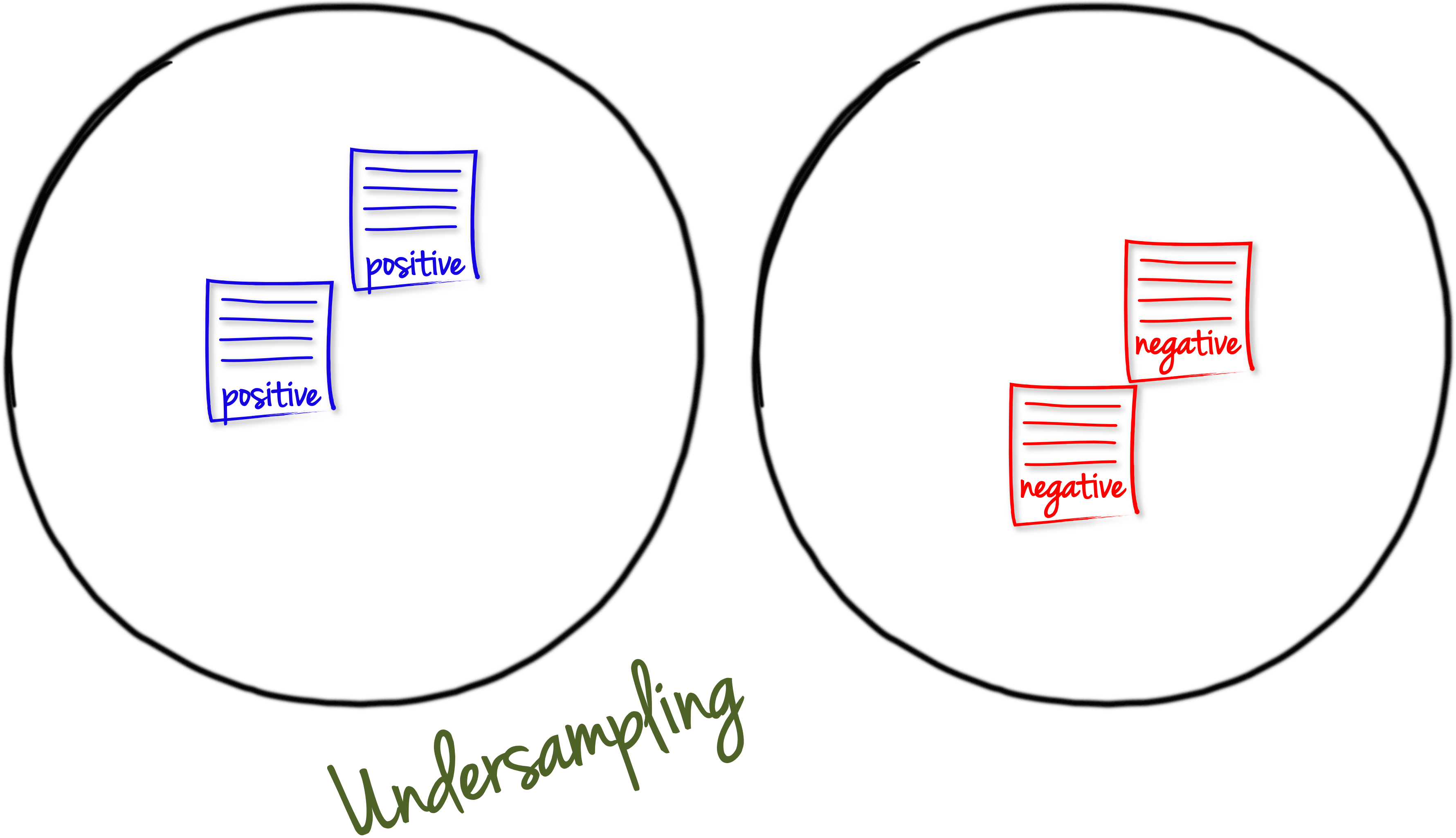
Performance Measures

Measure	Definition	References
A-Mean (AM)	$(\mathrm{TPR} + \mathrm{TNR})/2$	Chan & Stolfo (1998); Powers et al. (2005); Gu et al. (2009); KDD Cup 2001 challenge
G-Mean (GM)	$\sqrt{\mathrm{TPR} \cdot \mathrm{TNR}}$	Kubat & Matwin (1997); Daskalaki et al. (2006)
H-Mean (HM)	$2/(\frac{1}{\mathrm{TPR}} + \frac{1}{\mathrm{TNR}})$	Kennedy et al. (2009)
Q-Mean (QM)	$1 - ((\mathrm{FPR})^2 + (\mathrm{FNR})^2)/2$	Lawrence et al. (1998)
F_1	$2/(\frac{1}{\mathrm{Prec}} + \frac{1}{\mathrm{TPR}})$	Lewis & Gale (1994) Gu et al. (2009)
G-TP/PR	$\sqrt{\mathrm{TPR} \cdot \mathrm{Prec}}$	Daskalaki et al. (2006)
AUC-ROC	Area under ROC curve	Ling et al. (1998)
AUC-PR	Area under precision-recall curve	Davis & Goadrich (2006) Liu & Chawla (2011)

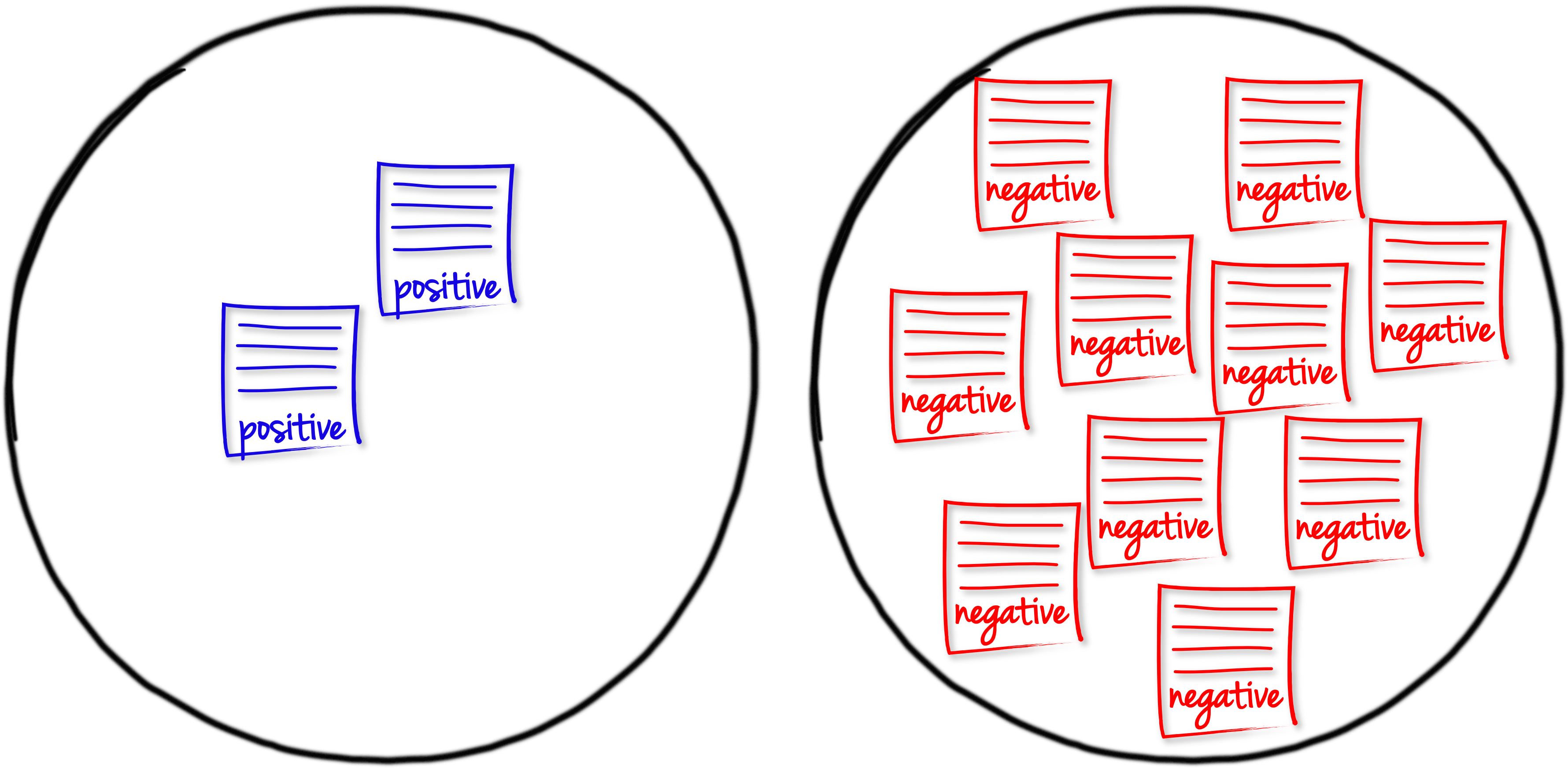
Class Imbalance



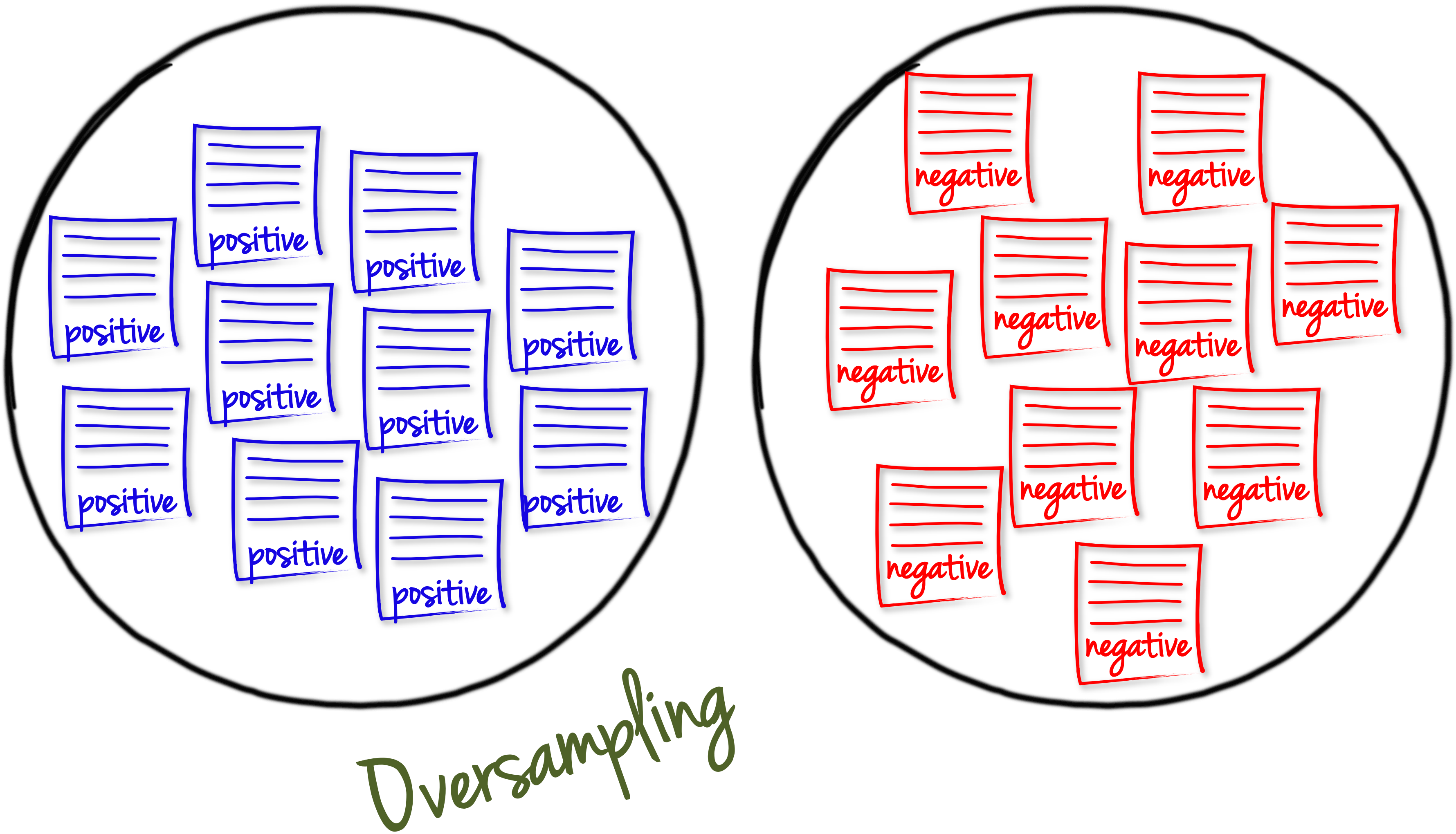
Class Imbalance



Class Imbalance



Class Imbalance

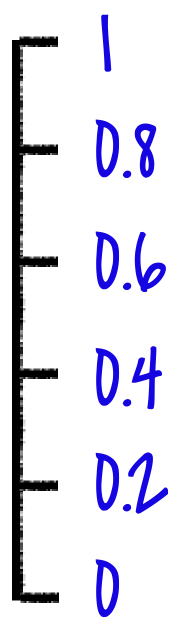


Class Imbalance

Plug-in Classifier

Learn a class probability estimator from S:

$$\widehat{\eta} : X \to [0, 1]$$

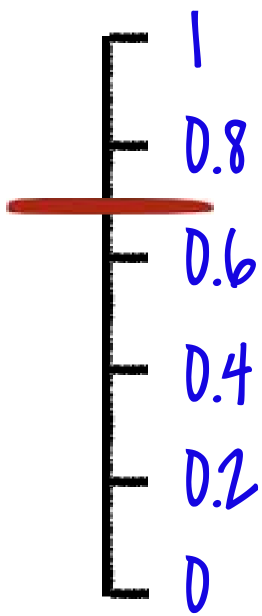


Class Imbalance

Plug-in Classifier

Learn a class probability estimator from S:

$$\widehat{\eta} : X \to [0, 1]$$



$\widehat{p}_S$

Empirical estimator of

$$p = \mathbf{P}(y = 1)$$

Class Imbalance

Balanced Empirical Risk Minimization

$$\frac{1}{\hat{p}S} \sum_{i:y_i=1} \ell(1, f(x_i)) + \frac{1}{1-\hat{p}S} \sum_{i:y_i=-1} \ell(-1, f(x_i))$$

Class Imbalance

Balanced Empirical Risk Minimization

$$\frac{1}{\hat{p}s}\sum_{i:y_i=1}\ell(1,f(x_i)) + \frac{1}{1-\hat{p}s}\sum_{i:y_i=-1}\ell(-1,f(x_i))$$

Empirical Balancing Terms

computed using empirical estimate of

$$p = \mathbf{P}(y=1)$$

Class Imbalance

Performance Measures

Measure	Definition	References
A-Mean (AM)	$(\mathrm{TPR} + \mathrm{TNR})/2$	Chan & Stolfo (1998); Powers et al. (2005); Gu et al. (2009); KDD Cup 2001 challenge
G-Mean (GM)	$\sqrt{\mathrm{TPR} \cdot \mathrm{TNR}}$	Kubat & Matwin (1997); Daskalaki et al. (2006)
H-Mean (HM)	$2/(\frac{1}{\mathrm{TPR}} + \frac{1}{\mathrm{TNR}})$	Kennedy et al. (2009)
Q-Mean (QM)	$1 - ((\mathrm{FPR})^2 + (\mathrm{FNR})^2)/2$	Lawrence et al. (1998)
F_1	$2/(\frac{1}{\mathrm{Prec}} + \frac{1}{\mathrm{TPR}})$	Lewis & Gale (1994) Gu et al. (2009)
G-TP/PR	$\sqrt{\mathrm{TPR} \cdot \mathrm{Prec}}$	Daskalaki et al. (2006)
AUC-ROC	Area under ROC curve	Ling et al. (1998)
AUC-PR	Area under precision-recall curve	Davis & Goadrich (2006) Liu & Chawla (2011)

Class Imbalance

AM-regret

$$\mathrm{regret}_D^{\mathrm{AM}}[h] = \sup_{h:\mathcal{X}\to\{\pm 1\}} \mathrm{AM}_D[h] \;-\; \mathrm{AM}_D[h]$$

AM-consistency

$$\mathrm{regret}_D^{\mathrm{AM}}[h_S] \xrightarrow{P} 0$$

Class Imbalance

AM-regret

$$\text{regret}_D^{\text{AM}}[h] = \sup_{h:\mathcal{X}\to\{\pm 1\}} \text{AM}_D[h] \; - \; \text{AM}_D[h]$$

AM-consistency

$$\text{regret}_D^{\text{AM}}[h_S] \xrightarrow{P} 0$$

Under mild conditions on the underlying distribution and under certain assumptions on the surrogate loss minimized, Plug-in approach and Balanced ERM are AM-consistent.

Menon, A., Narasimhan, H., Agarwal, S. and Chawla, S. "*On the statistical consistency of algorithms for binary classification under class imbalance* ", ICML 2013.

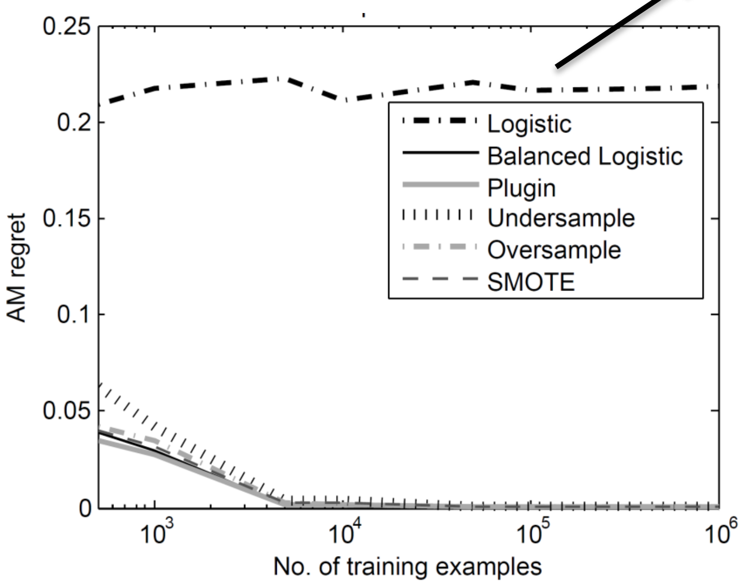
Class Imbalance

Loss	$\ell(y, f)$	Plug-in	Balanced ERM
Logistic	$\ln(1 + e^{-yf})$	✓	✓
Exponential	e^{-yf}	✓	✓
Square	$(1 - yf)^2$	✓	✓
Sq. Hinge	$((1 - yf)_+)^2$	✓	✓
Hinge	$(1 - yf)_+$	$\times$	✓

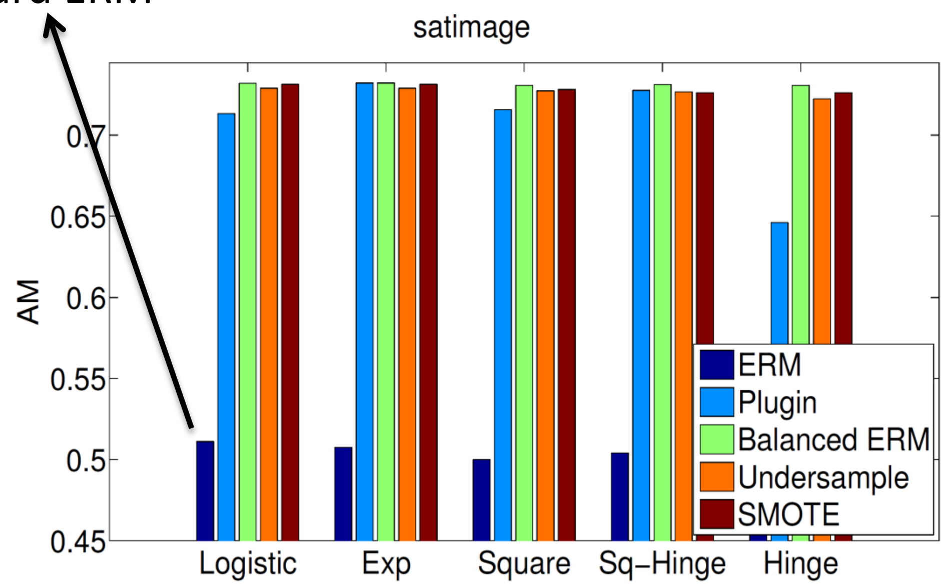
Menon, A., Narasimhan, H., Agarwal, S. and Chawla, S. "*On the statistical consistency of algorithms for binary classification under class imbalance* ", ICML 2013.

Class Imbalance

Standard ERM

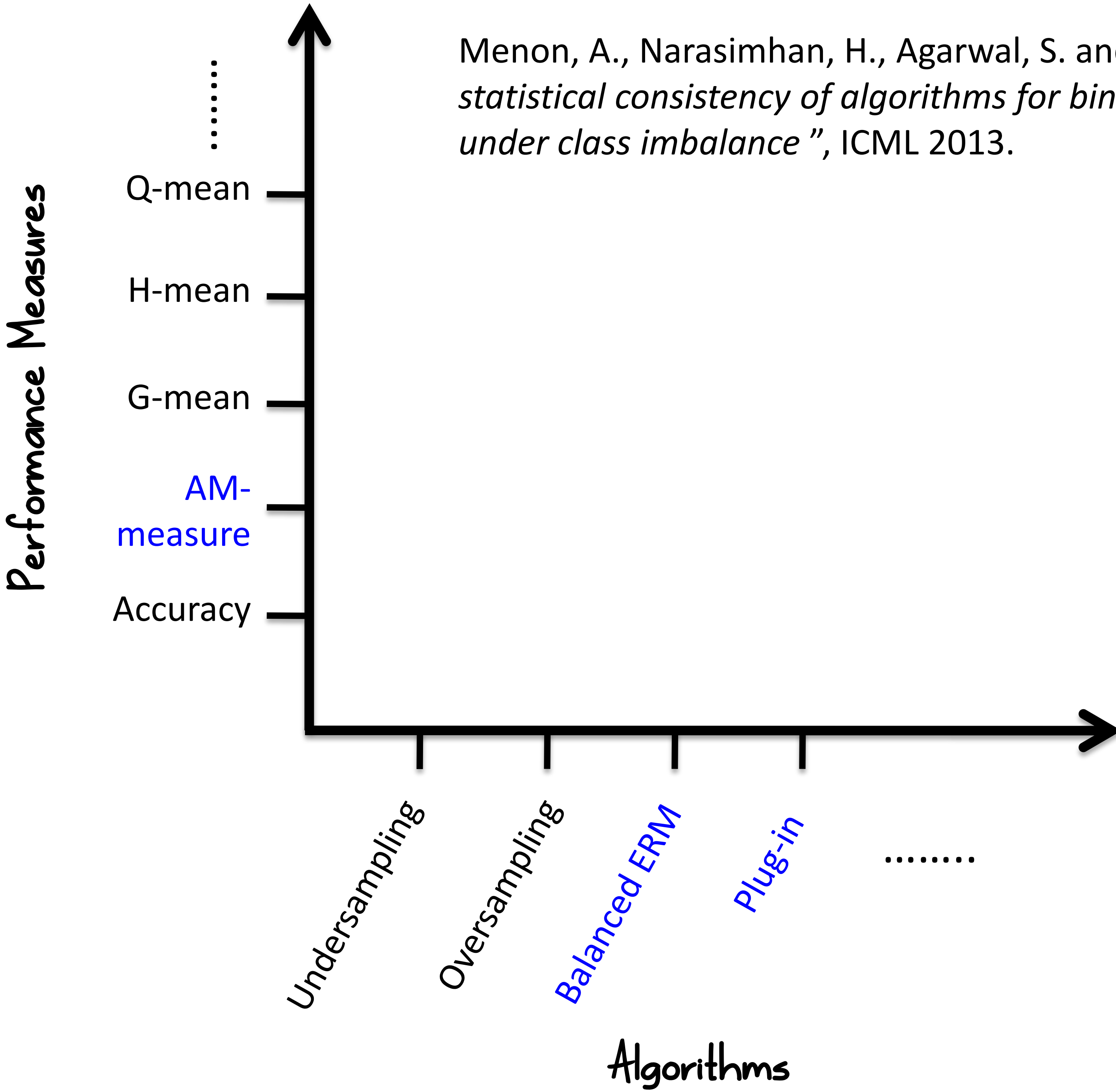


Synthetic data
p = 0.05

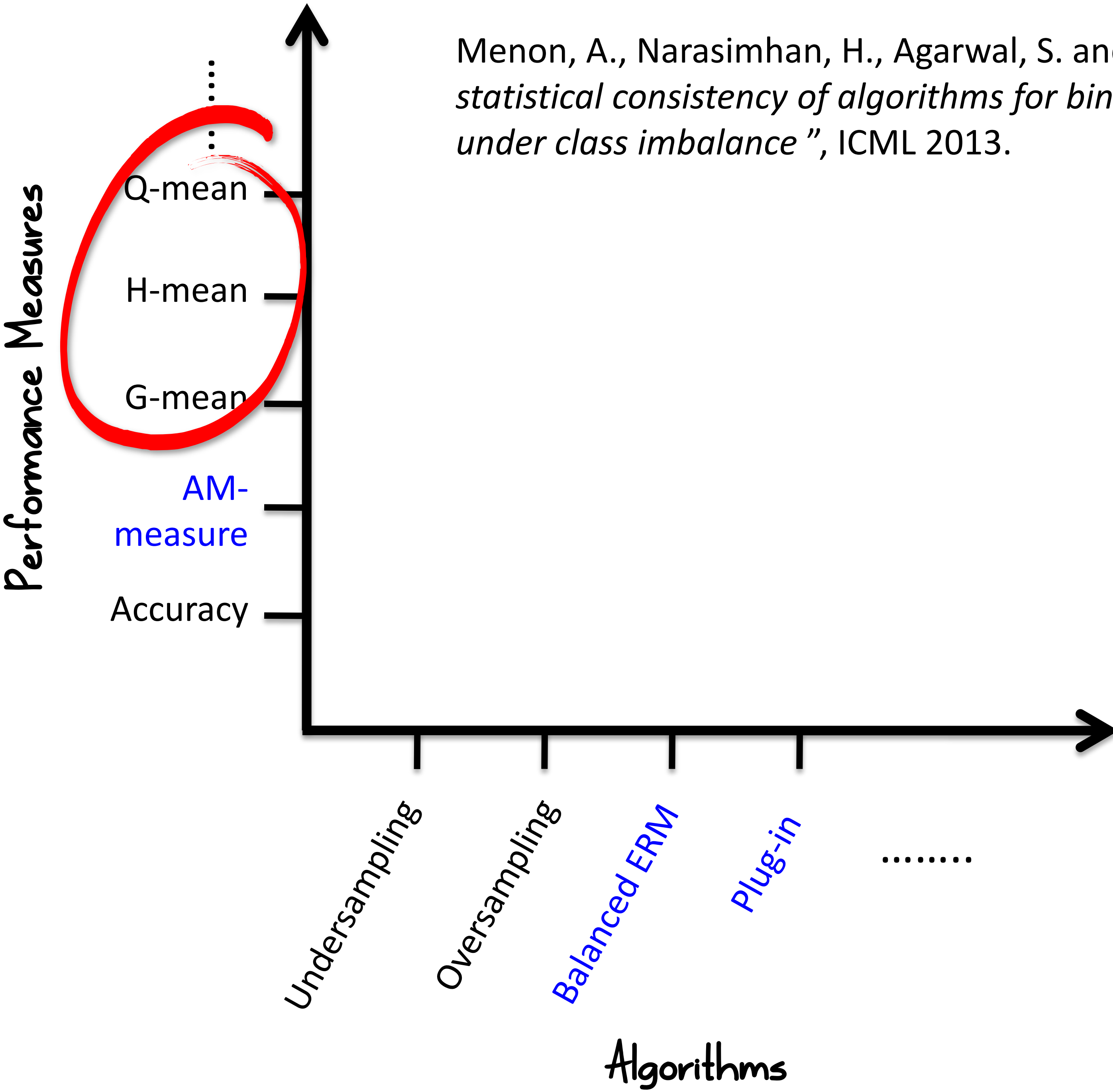


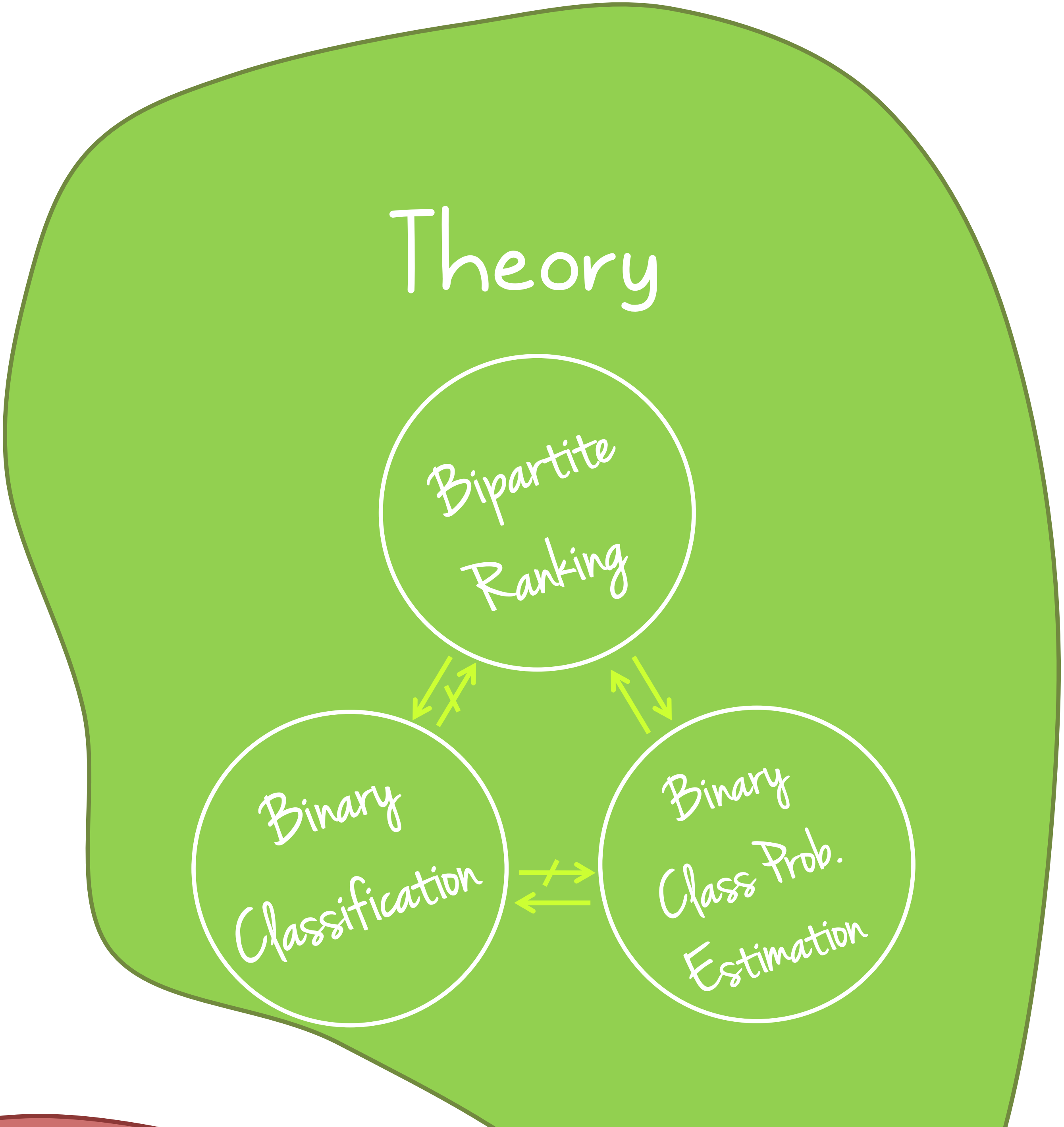
Real data
p = 0.097

Menon, A., Narasimhan, H., Agarwal, S. and Chawla, S. "*On the statistical consistency of algorithms for binary classification under class imbalance*", ICML 2013.



Menon, A., Narasimhan, H., Agarwal, S. and Chawla, S. "*On the statistical consistency of algorithms for binary classification under class imbalance*", ICML 2013.




Theory
Bipartite Ranking
Binary Classification
Binary Class Prob. Estimation

Binary Classification

$$h : X \to \{\pm 1\}$$

E.g. Spam Classification

allspammedup.com

Binary Classification

Bipartite Ranking

$$h : X \to \{\pm 1\}$$

$$f : X \to \mathbb{R}$$

E.g. Spam Classification

E.g. Information Retrieval

allspammedup.com

fusionsedge.com

Binary Classification

$h : X \to \{\pm 1\}$

E.g. Spam Classification

allspammedup.com

Bipartite Ranking

$f : X \to \mathbb{R}$

E.g. Information Retrieval

fusionsedge.com

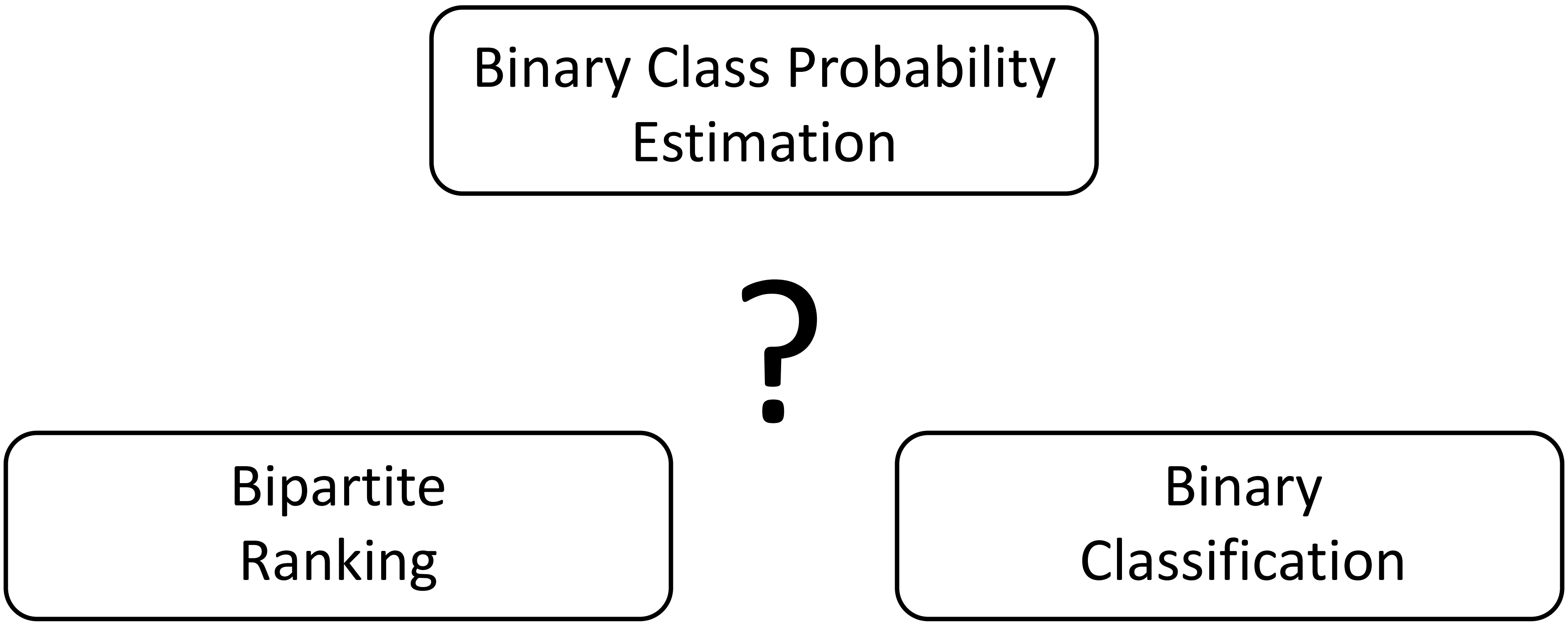
Binary Class Prob. Estimation

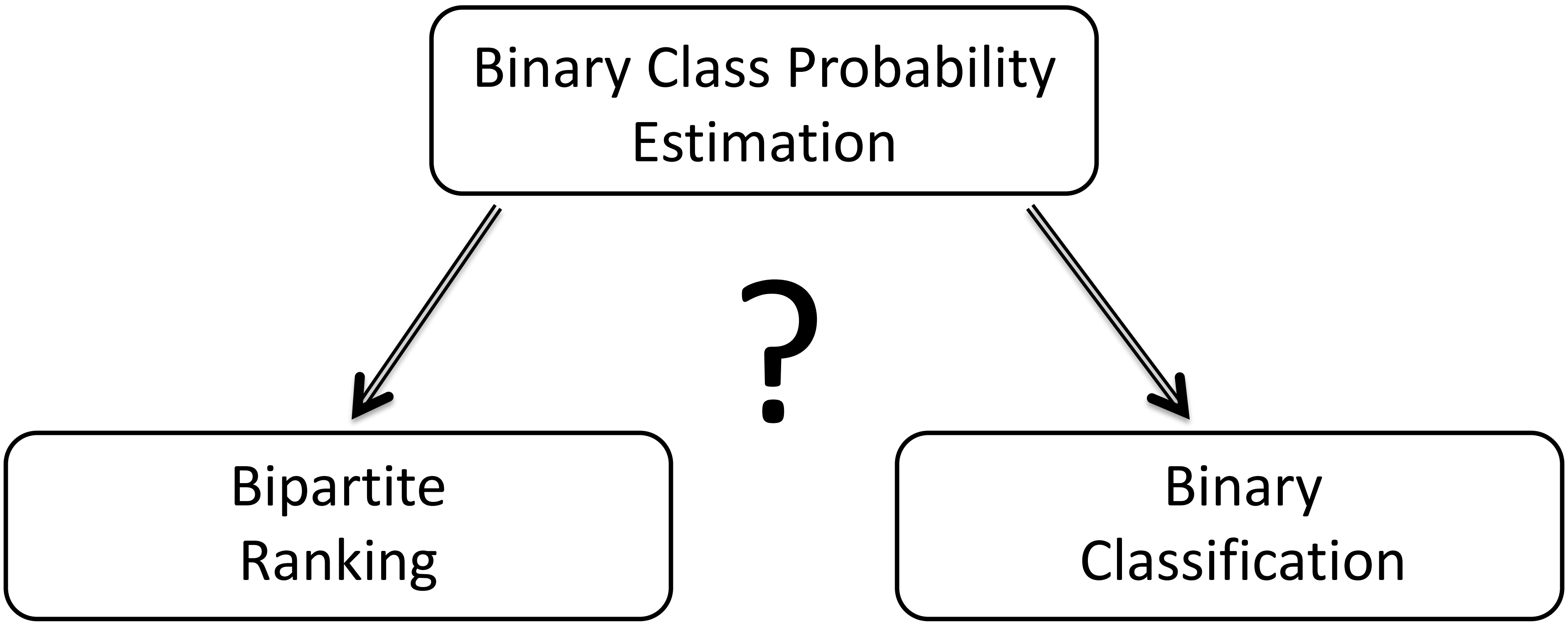
$\widehat{\eta} : X \to [0, 1]$

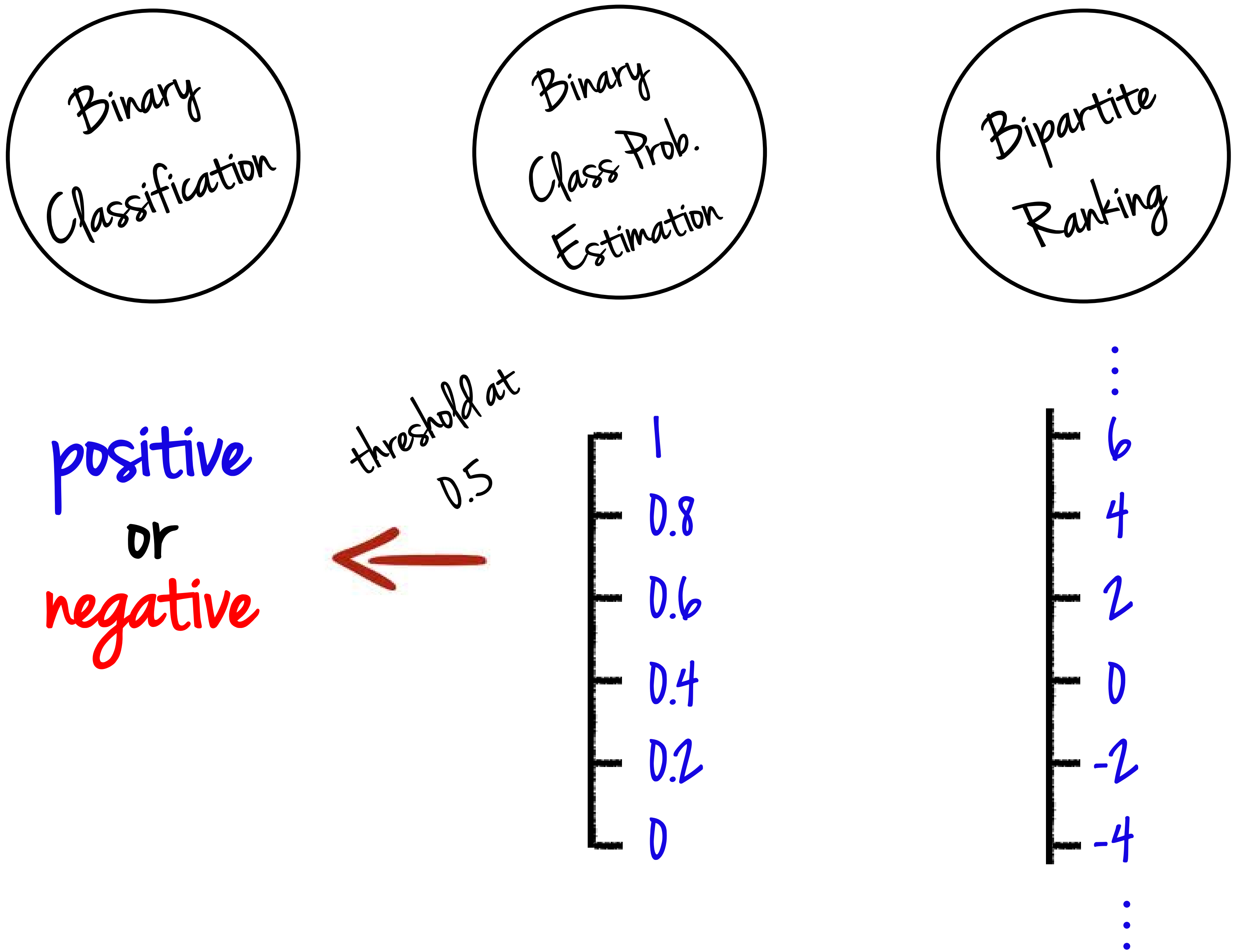
E.g. Click-through Rates

Probability of click?

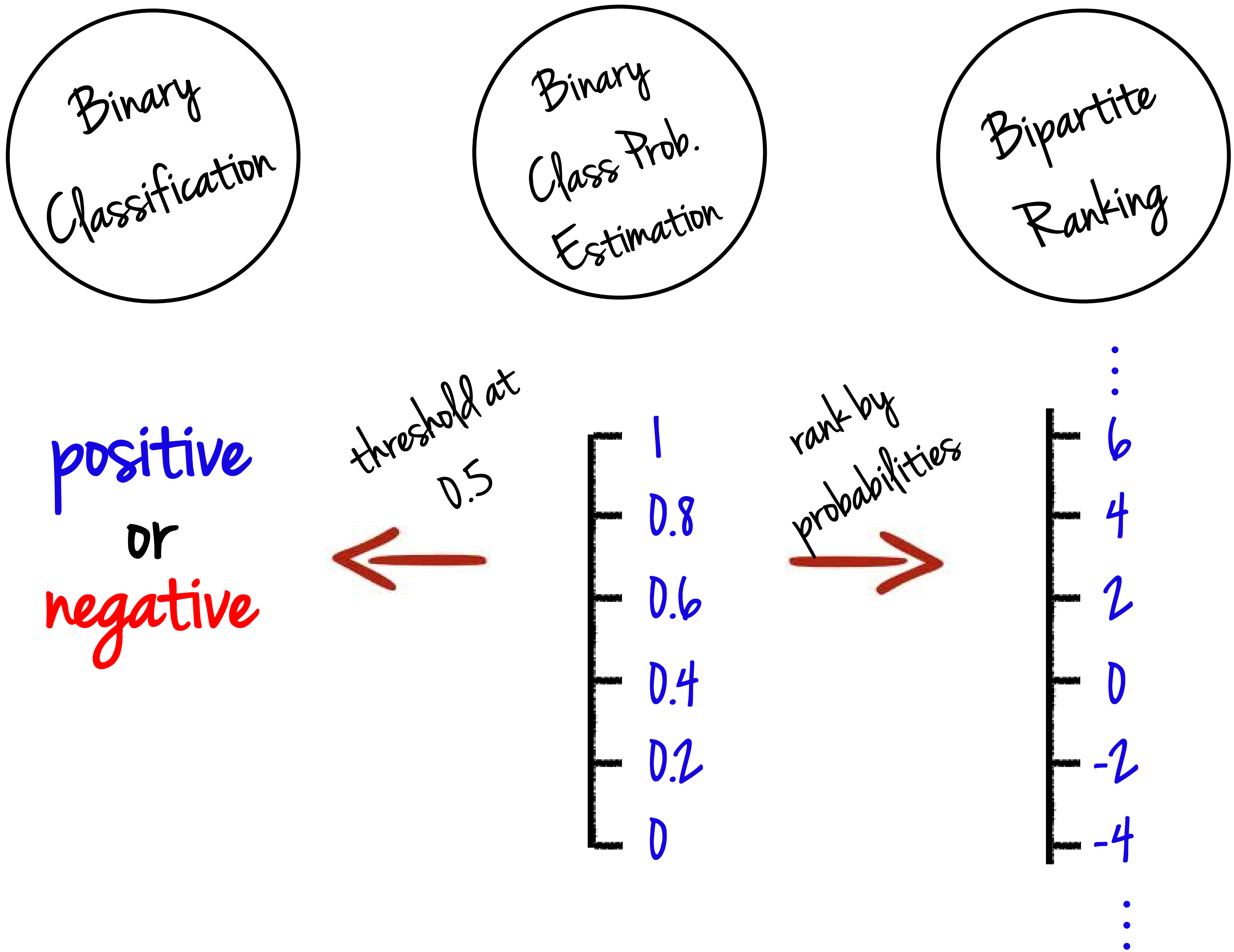
optimum7.com


Binary Class Probability Estimation
?
Bipartite Ranking
Binary Classification

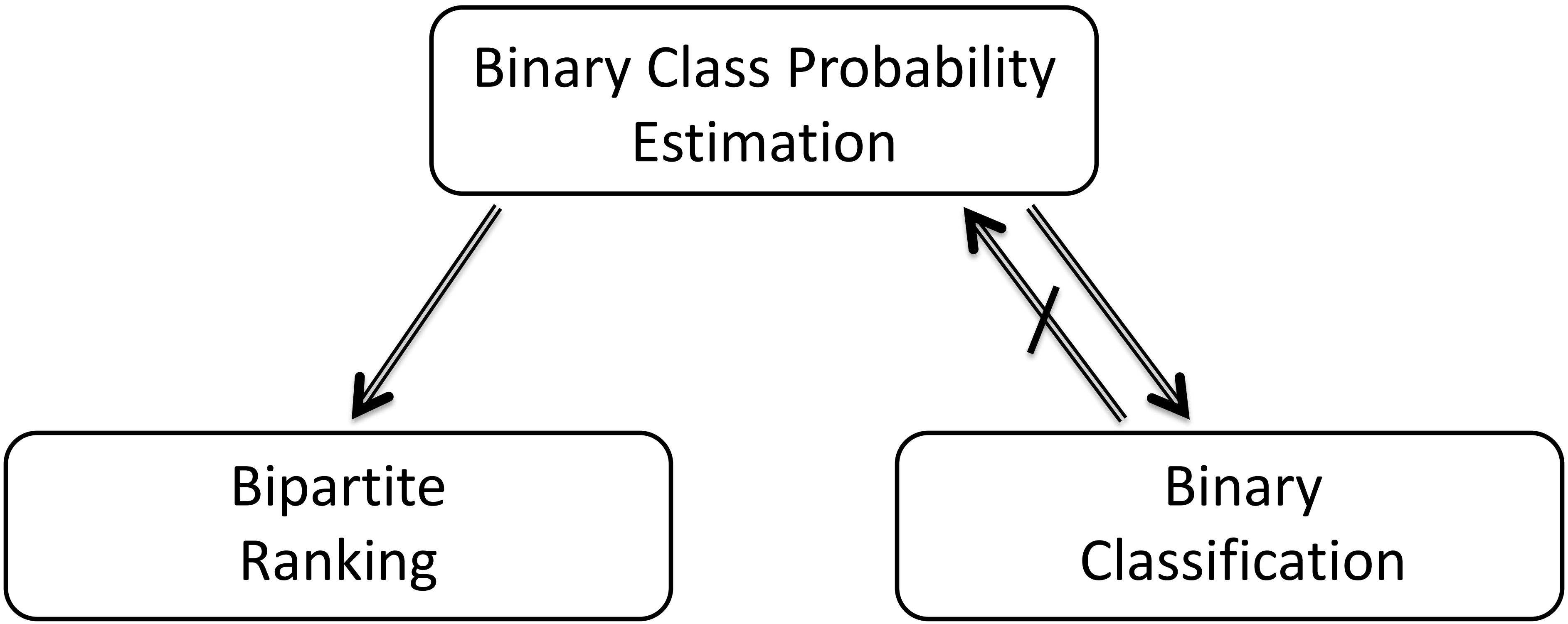

Binary Class Probability
Estimation
?
Bipartite
Ranking
Binary
Classification

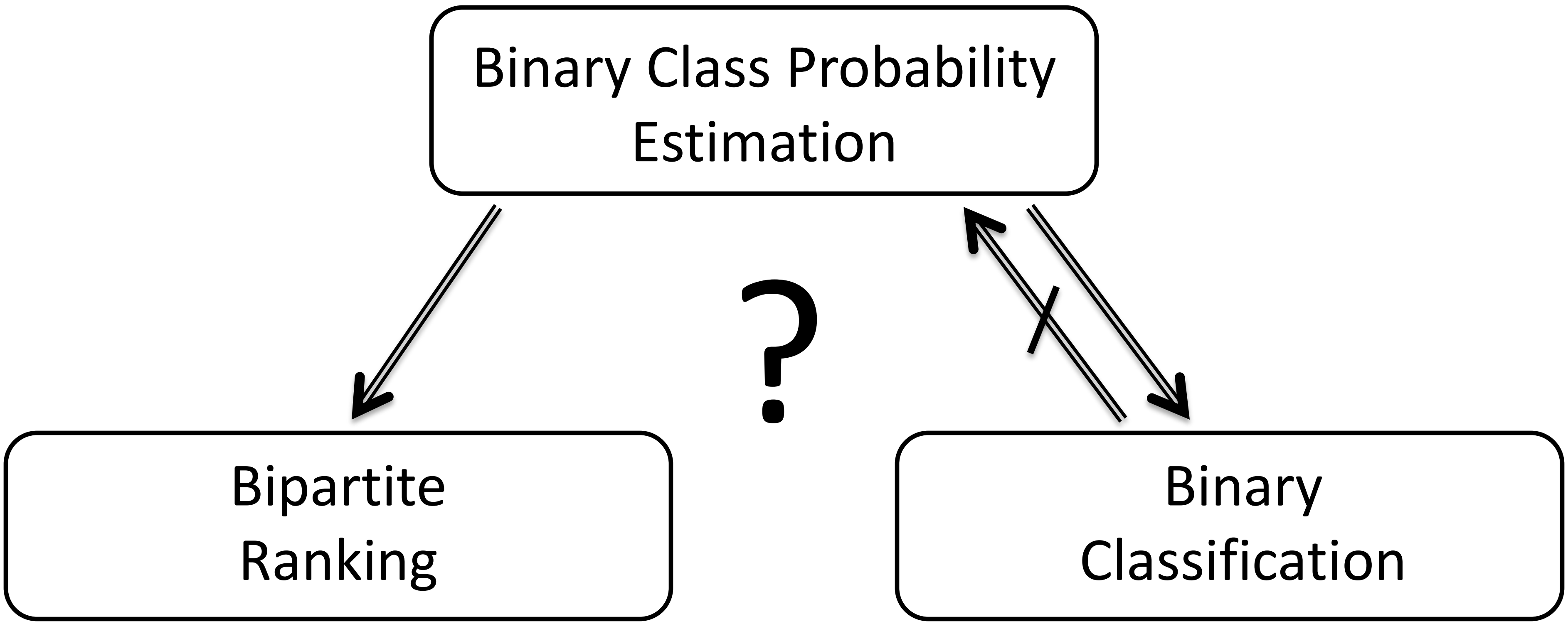


Binary Classification
Binary Class Prob. Estimation
Bipartite Ranking
positive
or
negative
threshold at 0.5
1
0.8
0.6
0.4
0.2
0
6
4
2
0
-2
-4

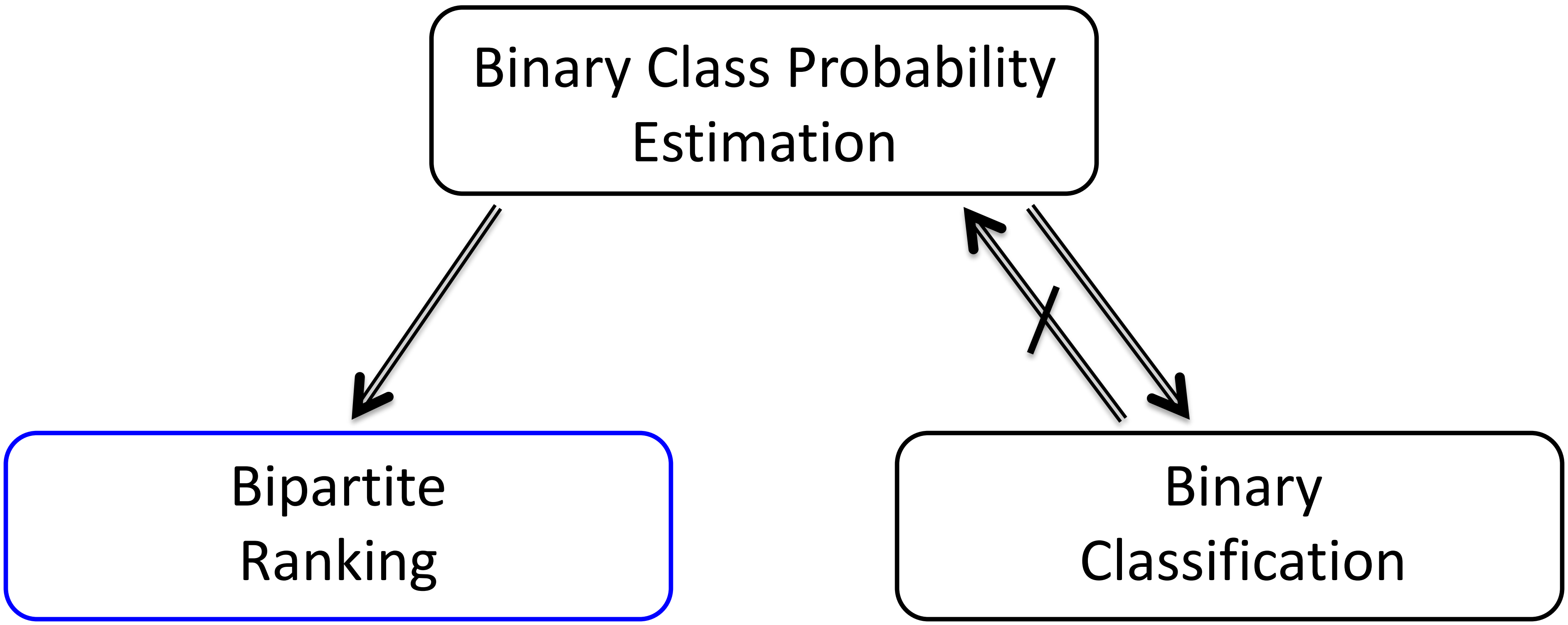


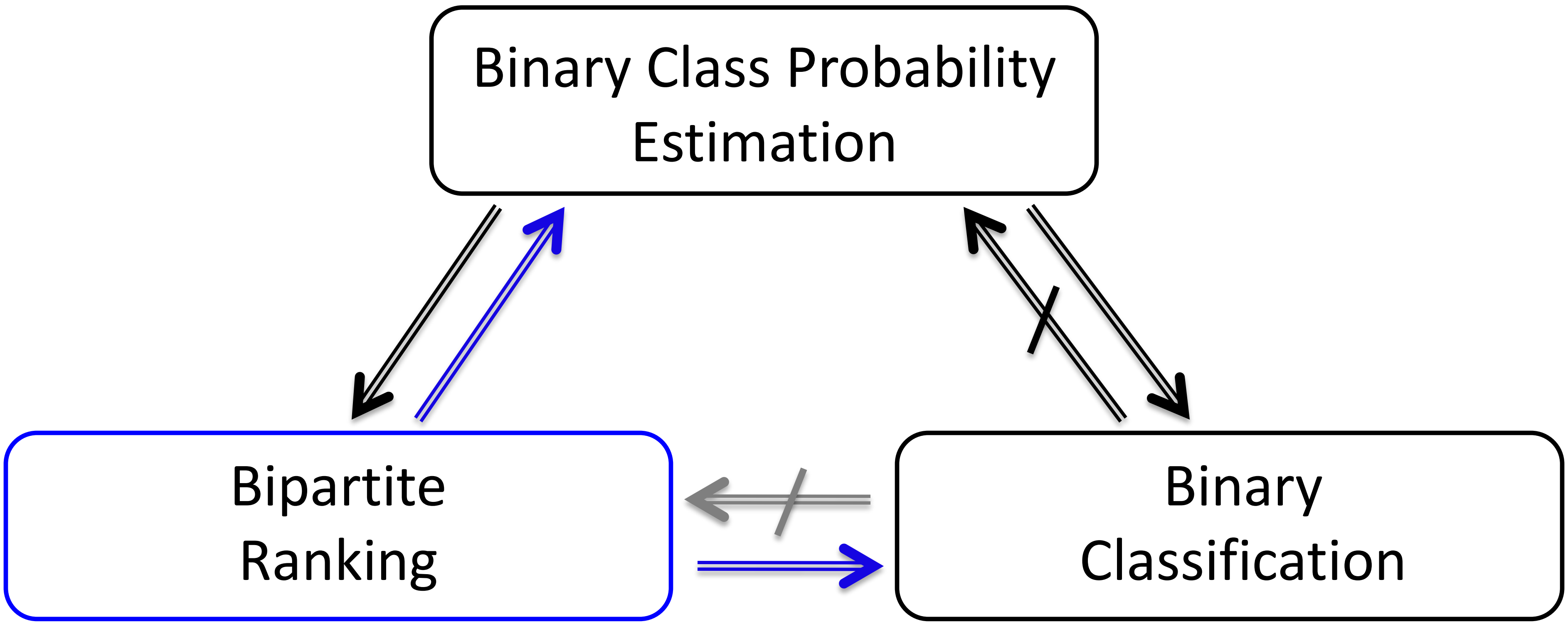
Binary Classification
Binary Class Prob. Estimation
Bipartite Ranking
positive
or
negative
threshold at 0.5
rank by probabilities
1
0.8
0.6
0.4
0.2
0
6
4
2
0
-2
-4


Binary Class Probability
Estimation
Bipartite
Ranking
Binary
Classification



Binary Class Probability Estimation
?
Bipartite Ranking
Binary Classification


Binary Class Probability Estimation
Bipartite Ranking
Binary Classification



Narasimhan, H. and Agarwal, S. "*On the relationship between binary classification, bipartite ranking, and binary class probability estimation*". In NIPS 2013. To appear.

Binary Classification

$$h : X \to \{\pm 1\}$$

Binary Classification

$$h : X \to \{\pm 1\}$$

$$\mathrm{er}_D^{0\text{-}1}[h] \;=\; \mathbf{E}_{(x,y)\sim D}\big[\mathbf{1}(h(x) \neq y)\big]$$

$$\mathrm{er}_D^{0\text{-}1,*} \;=\; \inf_{h:X\to\{\pm 1\}} \mathrm{er}_D^{0\text{-}1}[h]$$

Binary Classification

$$h : X \to \{\pm 1\}$$

$$\mathrm{er}_D^{0\text{-}1}[h] \;=\; \mathbf{E}_{(x,y)\sim D}\big[\mathbf{1}(h(x) \neq y)\big]$$

$$\mathrm{er}_D^{0\text{-}1,*} \;=\; \inf_{h:X\to\{\pm 1\}} \mathrm{er}_D^{0\text{-}1}[h]$$

$$\boxed{\mathrm{regret}_D^{0\text{-}1}[h] \;=\; \mathrm{er}_D^{0\text{-}1}[h] - \mathrm{er}_D^{0\text{-}1,*}}$$

Bipartite Ranking

$$f : X \to \mathbb{R}$$

Bipartite Ranking

$$f : X \to \mathbb{R}$$

$$\mathrm{er}_D^{\mathrm{rank}}[f] = \mathbf{E}\Big[\mathbf{1}\big((y - y')(f(x) - f(x')) < 0\big) \,\big|\, y \neq y'\Big]$$

Bipartite Ranking

$$f : X \to \mathbb{R}$$

$$\mathrm{er}_D^{\mathrm{rank}}[f] = \mathbf{E}\Big[\mathbf{1}\big((y - y')(f(x) - f(x')) < 0\big) \,\big|\, y \neq y'\Big]$$

$$\mathrm{er}_D^{\mathrm{rank},*} = \inf_{f : X \to \mathbb{R}} \mathrm{er}_D^{\mathrm{rank}}[f]$$

Bipartite Ranking

$$f : X \to \mathbb{R}$$

$$\mathrm{er}_D^{\mathrm{rank}}[f] = \mathbf{E}\left[\mathbf{1}\Big((y - y')(f(x) - f(x')) < 0\Big) \,\Big|\, y \neq y'\right]$$

$$\mathrm{er}_D^{\mathrm{rank},*} = \inf_{f:X\to\mathbb{R}} \mathrm{er}_D^{\mathrm{rank}}[f]$$

$$\boxed{\mathrm{regret}_D^{\mathrm{rank}}[f] \; = \; \mathrm{er}_D^{\mathrm{rank}}[f] - \mathrm{er}_D^{\mathrm{rank},*}}$$

Binary Class Probability Estimation

$$\widehat{\eta} : X \to [0, 1]$$

$$\eta(x) = \mathbf{P}(y = 1 \mid x)$$

Binary Class Probability Estimation

$$\widehat{\eta} : X \to [0, 1]$$

$$\eta(x) = \mathbf{P}(y = 1 \mid x)$$

$$\boxed{\mathrm{regret}^{\mathrm{sq}}_{D}[\,\widehat{\eta}\,] \;=\; \mathbf{E}_x\Big[\big(\widehat{\eta}(x) - \eta(x)\big)^2\Big]}$$

$$S = ((x_1, y_1), \ldots, (x_n, y_n)) \in (X \times \{\pm 1\})^n$$

drawn iid from D

0-1

Consistency

$$\boxed{\mathrm{regret}_D^{0\text{-}1}[h_S] \xrightarrow{P} 0}$$

$$S = ((x_1, y_1), \ldots, (x_n, y_n)) \in (X \times \{\pm 1\})^n$$
drawn iid from D

0-1 Consistency

$$\boxed{\mathrm{regret}_D^{0\text{-}1}[h_S] \xrightarrow{P} 0}$$

Ranking Consistency

$$\boxed{\mathrm{regret}_D^{\mathrm{rank}}[f_S] \xrightarrow{P} 0}$$

$$S = ((x_1, y_1), \ldots, (x_n, y_n)) \in (X \times \{\pm 1\})^n$$

drawn iid from D

0-1 Consistency

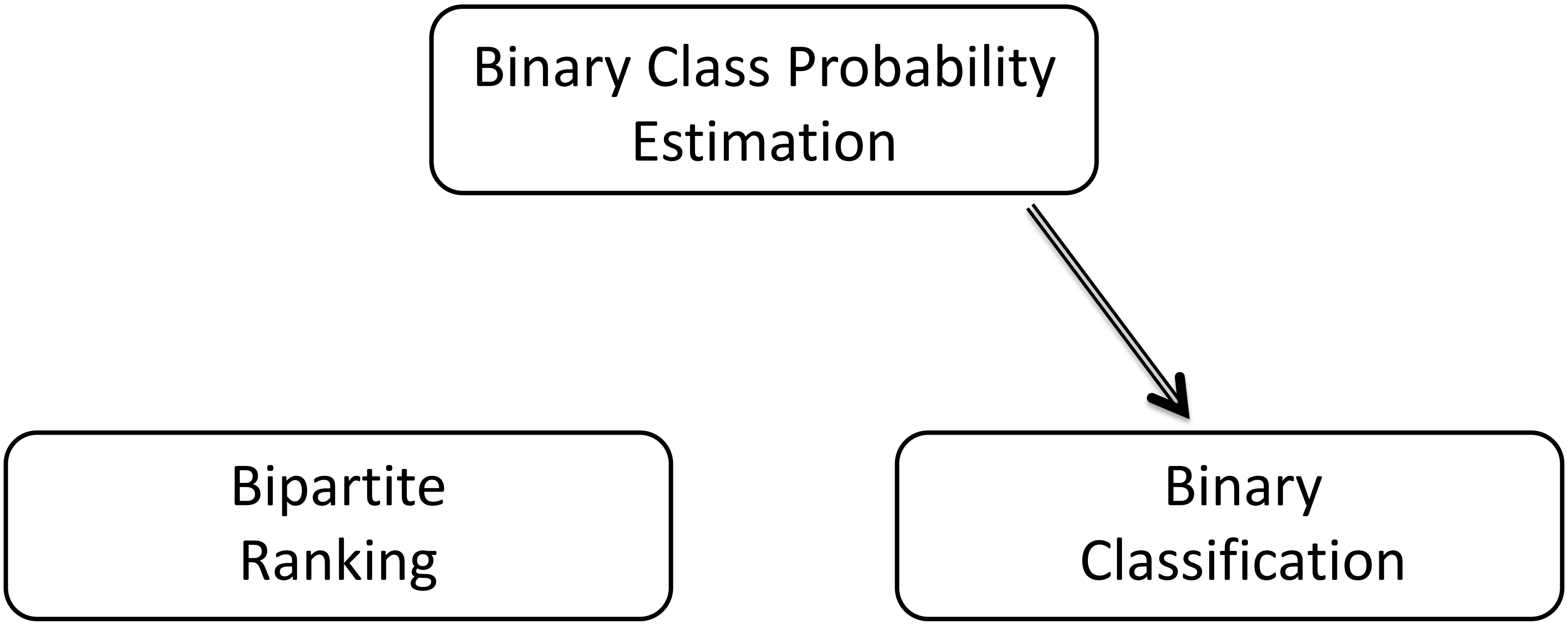
$$\boxed{\mathrm{regret}_D^{0\text{-}1}[h_S] \xrightarrow{P} 0}$$

Ranking Consistency

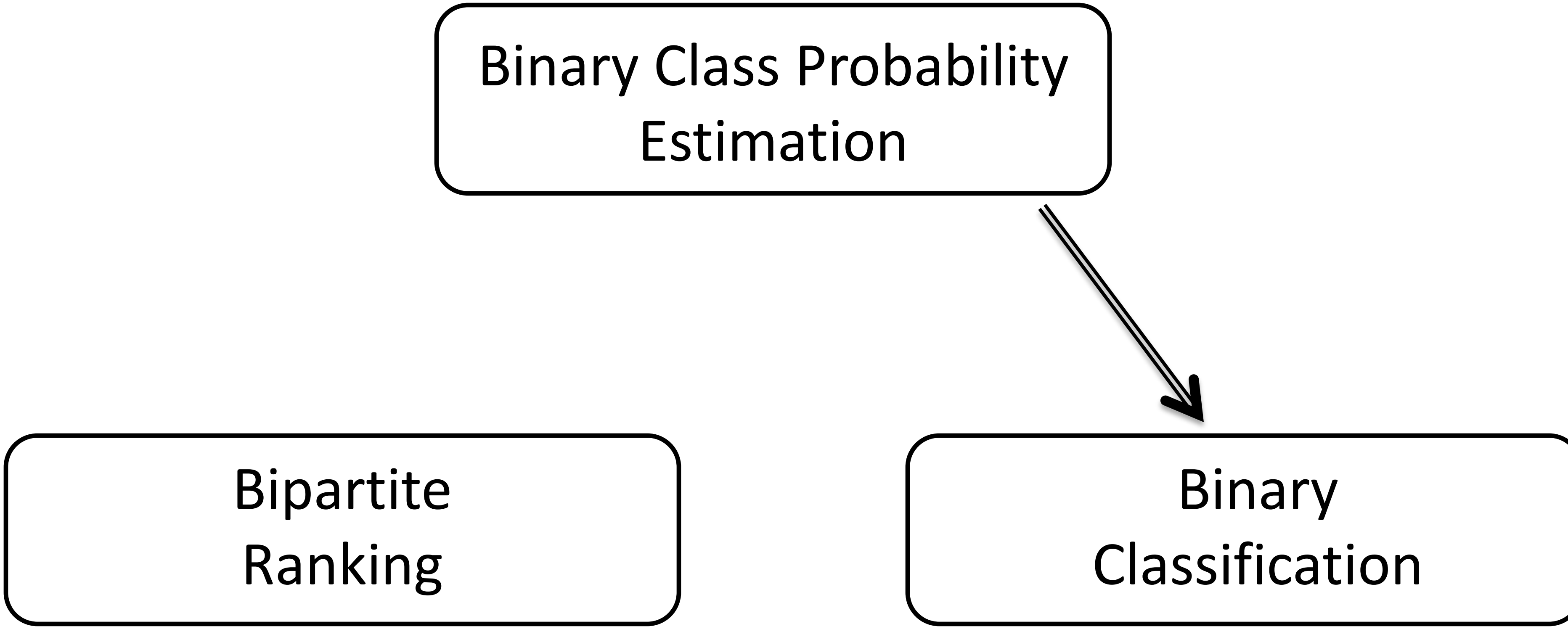
$$\boxed{\mathrm{regret}_D^{\mathrm{rank}}[f_S] \xrightarrow{P} 0}$$

CPE Consistency

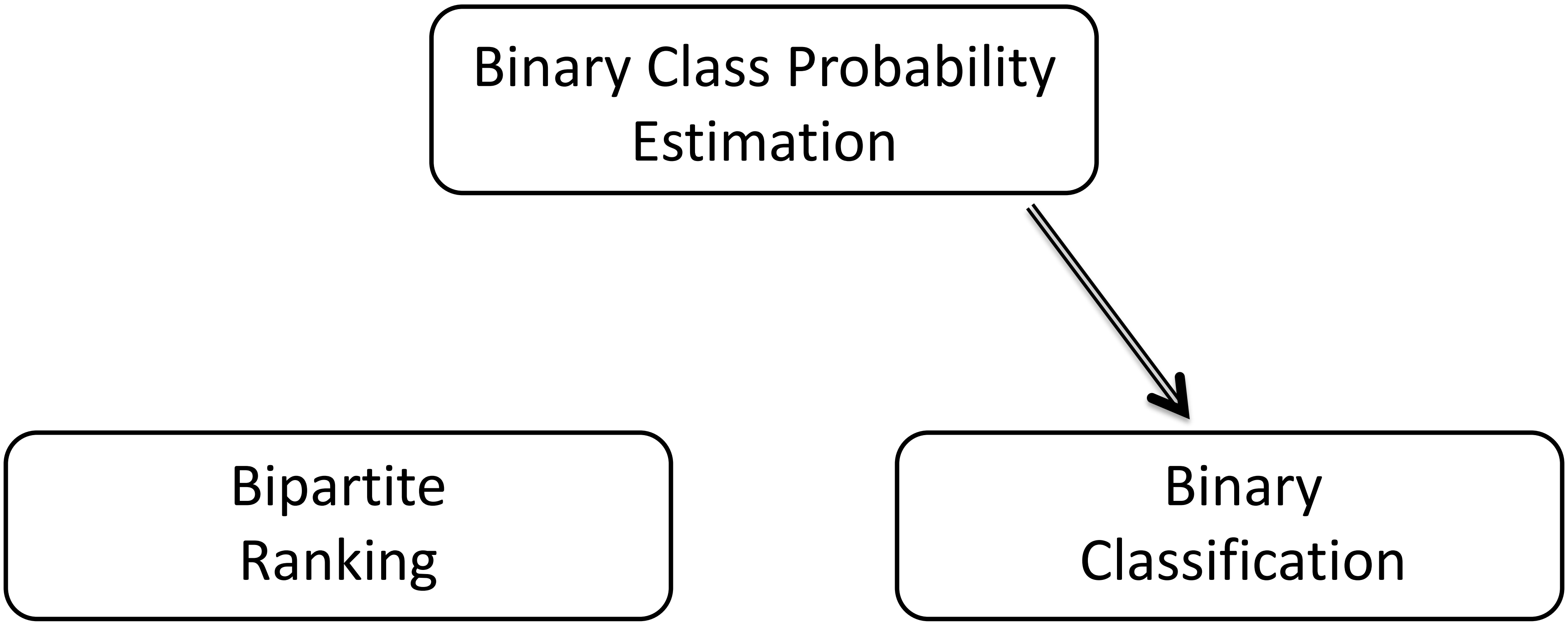
$$\boxed{\mathrm{regret}_D^{\mathrm{sq}}[\hat{\eta}_S] \xrightarrow{P} 0}$$



$$\mathrm{regret}_D^{0\text{-}1}[\mathrm{sign}\circ(\widehat{\eta}-\tfrac{1}{2})] \;\leq\; \sqrt{\mathrm{regret}_D^{\mathrm{sq}}[\,\widehat{\eta}\,]}$$



$$\mathrm{regret}_D^{0\text{-}1}\left[\mathrm{sign}\circ\left(\widehat{\eta}-\tfrac{1}{2}\right)\right] \;\leq\; \boxed{\sqrt{\mathrm{regret}_D^{\mathrm{sq}}[\,\widehat{\eta}\,]}} \;\xrightarrow{P}\; 0$$



$$\boxed{\mathrm{regret}_D^{0\text{-}1}\left[\mathrm{sign}\circ\left(\widehat{\eta}-\tfrac{1}{2}\right)\right]} \;\le\; \boxed{\sqrt{\mathrm{regret}_D^{\mathrm{sq}}[\,\widehat{\eta}\,]}}$$

$$\xrightarrow{P} 0 \qquad\qquad \xrightarrow{P} 0$$

Binary Class Probability Estimation

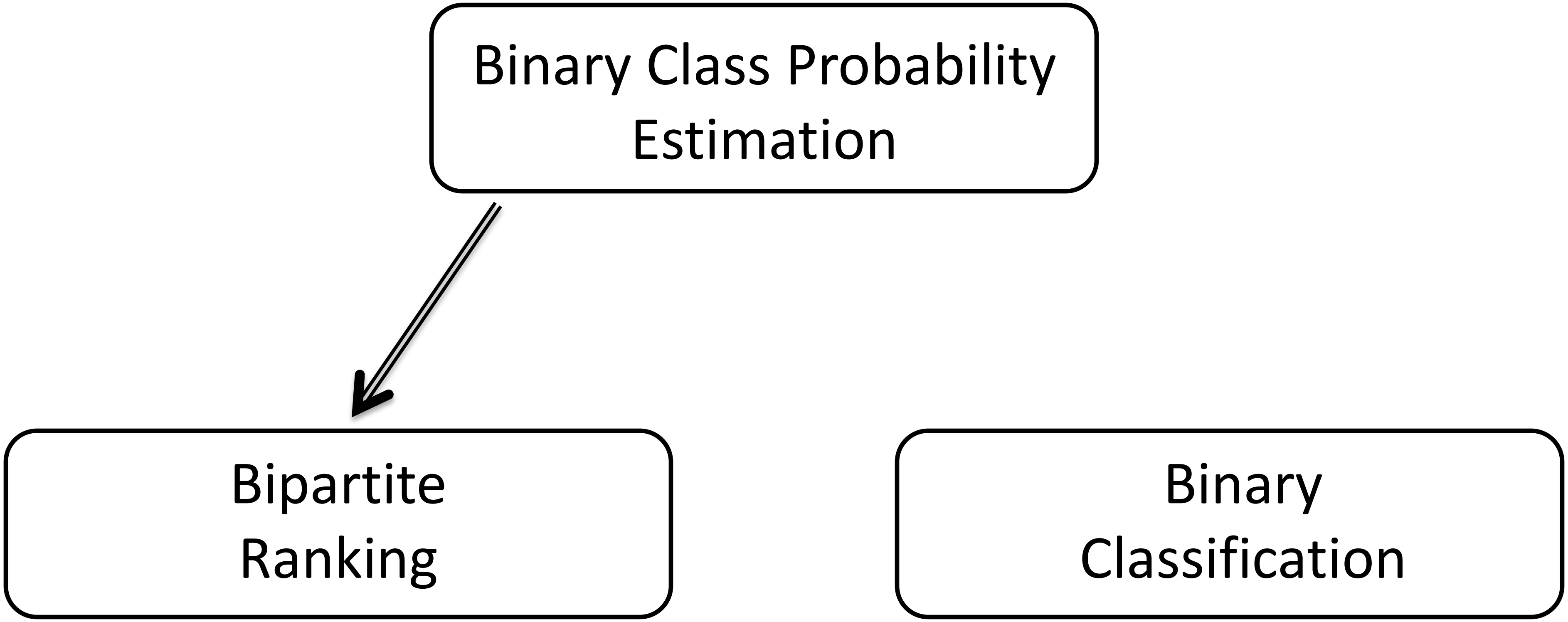
Bipartite Ranking

Binary Classification

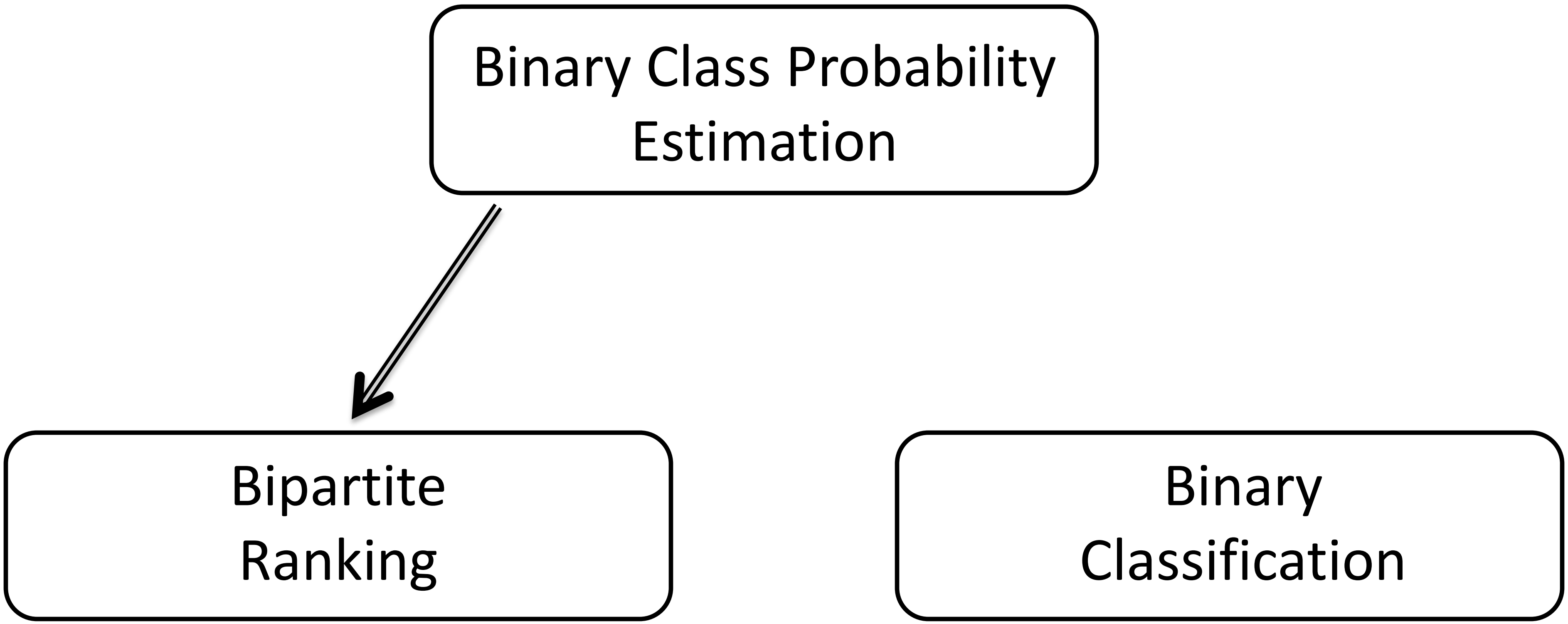
Regret Transfer Bound

$$\boxed{\mathrm{regret}_D^{0\text{-}1}\left[\mathrm{sign}\circ\left(\widehat{\eta}-\tfrac{1}{2}\right)\right]} \le \boxed{\sqrt{\mathrm{regret}_D^{\mathrm{sq}}[\,\widehat{\eta}\,]}}$$

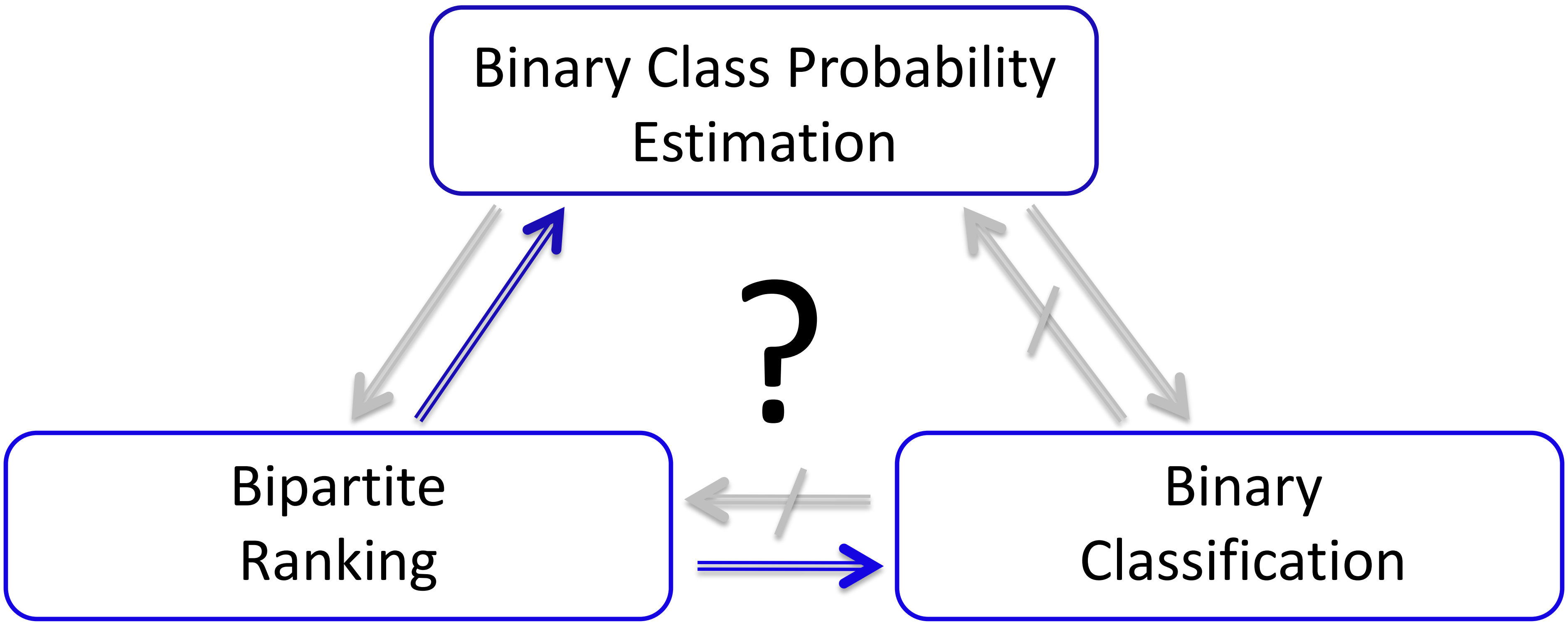
$$\xrightarrow{P} 0 \qquad\qquad \xrightarrow{P} 0$$

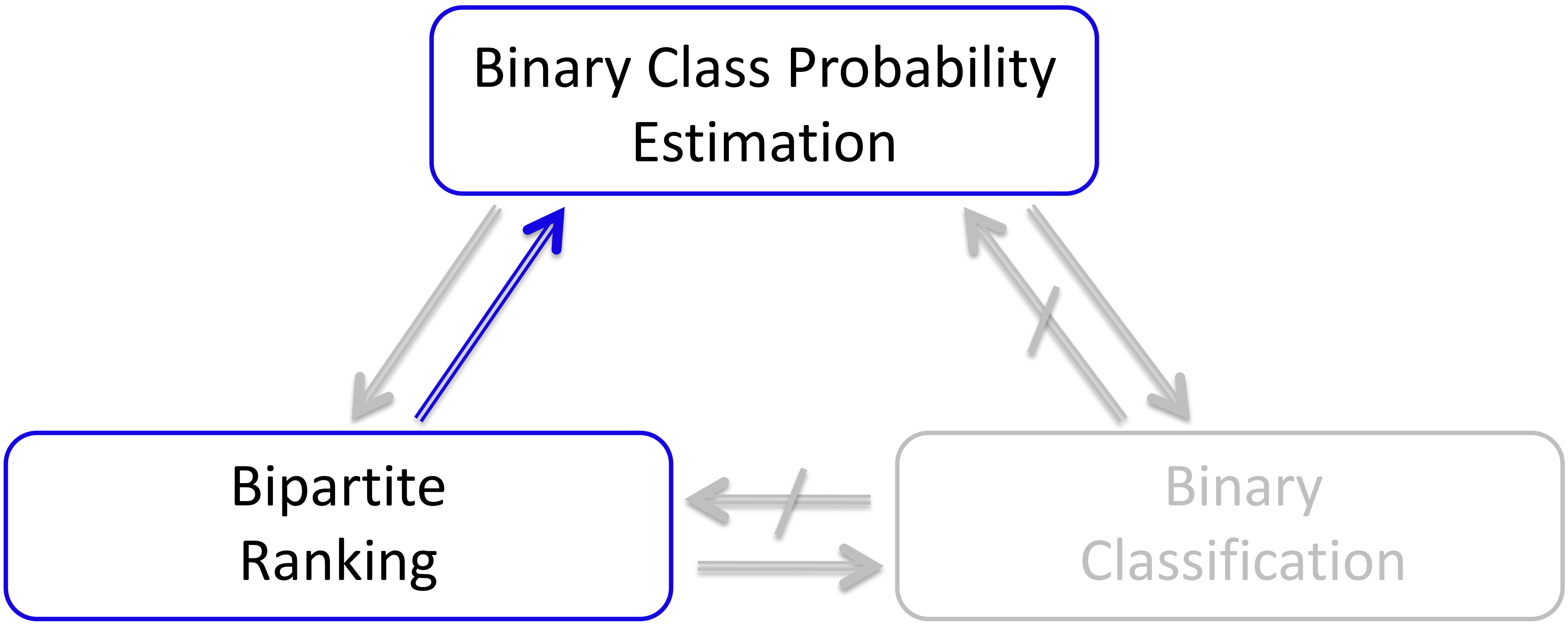


$$\mathrm{regret}_D^{\mathrm{rank}}[\,\widehat{\eta}\,] \;\le\; \frac{1}{p(1-p)}\sqrt{\mathrm{regret}_D^{\mathrm{sq}}[\,\widehat{\eta}\,]}$$



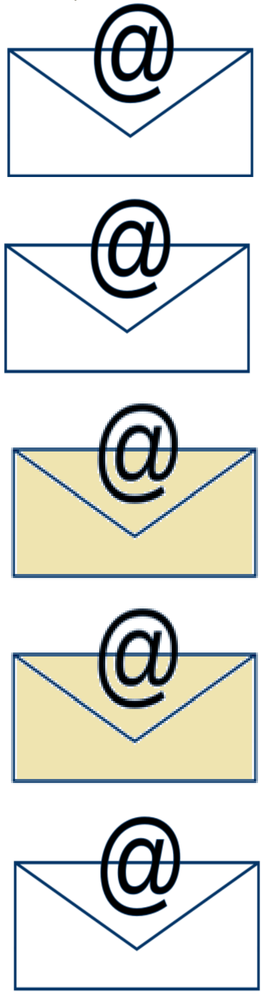
$$\underset{\xrightarrow{P}\ 0}{\boxed{\mathrm{regret}_D^{\mathrm{rank}}[\,\widehat{\eta}\,]}} \leq \frac{1}{p(1-p)}\sqrt{\underset{\xrightarrow{P}\ 0}{\boxed{\mathrm{regret}_D^{\mathrm{sq}}[\,\widehat{\eta}\,]}}}$$


Binary Class Probability Estimation
?
Bipartite Ranking
Binary Classification

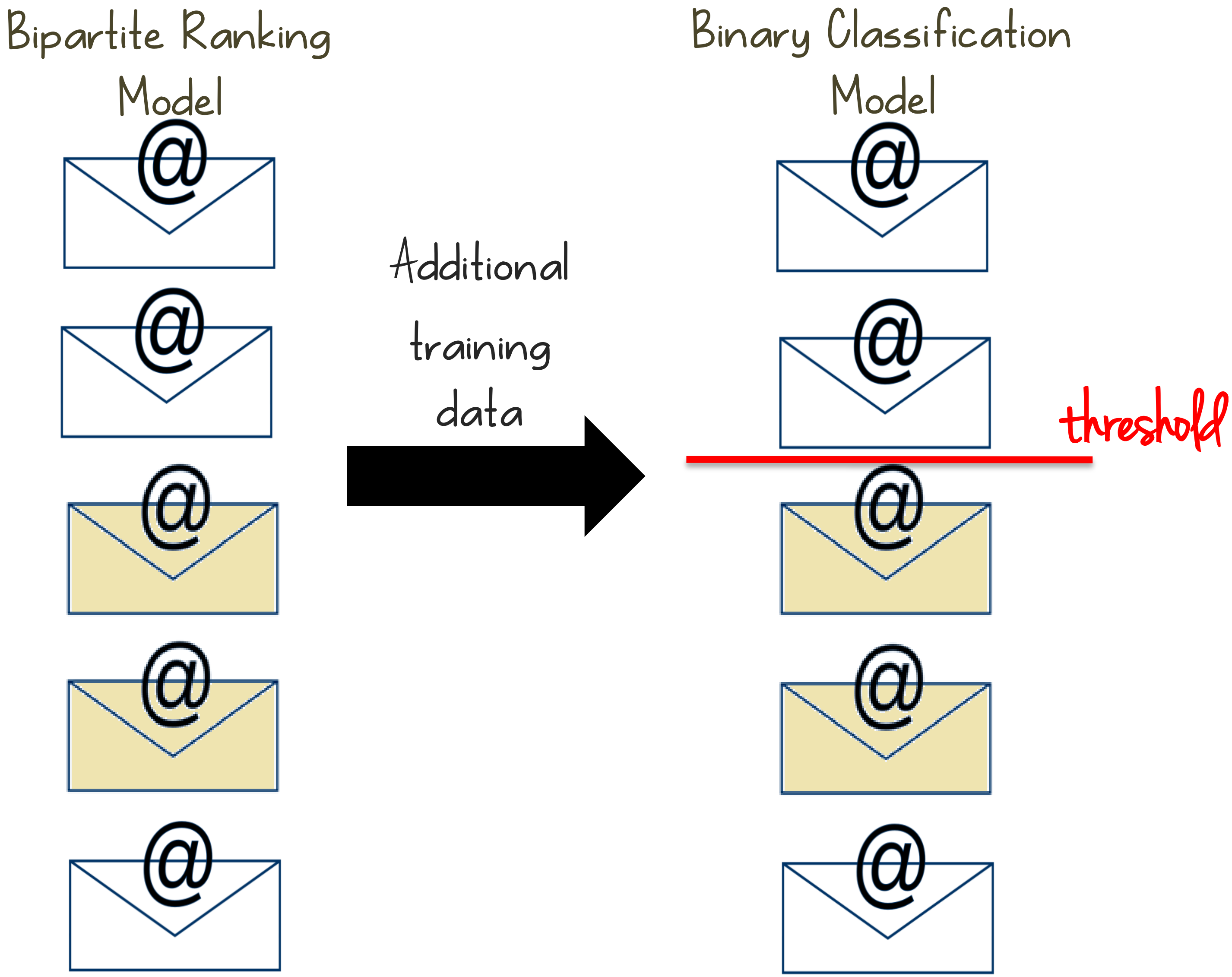

Binary Class Probability Estimation
Bipartite Ranking
Binary Classification

Bipartite Ranking → Binary Classification

Bipartite Ranking Model



Bipartite Ranking → Binary Classification



Bipartite Ranking → Binary Classification

$$f : X \to \mathbb{R}$$

$$t^*_{D,f} \in \operatorname{argmin}_{t \in [-\infty,\infty]} \left\{ \mathrm{er}^{0\text{-}1}_D \left[\mathrm{sign} \circ \left(f - t \right) \right] \right\}$$

Bipartite Ranking → Binary Classification

$$f : X \to \mathbb{R}$$

$$t^*_{D,f} \in \operatorname{argmin}_{t \in [-\infty,\infty]} \left\{ \mathrm{er}^{0\text{-}1}_D [\mathrm{sign} \circ (f - t)] \right\}$$

Idealized (Weak) Regret Transfer Bound

$$\boxed{\mathrm{regret}^{0\text{-}1}_D [\mathrm{sign} \circ (f - t^*_{D,f})] \;\le\; \sqrt{2p(1-p)\, \mathrm{regret}^{\mathrm{rank}}_D [f]}}$$

Bipartite Ranking → Binary Classification

$$f : X \to \mathbb{R}$$

$$t^*_{D,f} \in \operatorname{argmin}_{t \in [-\infty,\infty]} \left\{ \mathrm{er}^{0\text{-}1}_D [\mathrm{sign} \circ (f - t)] \right\}$$

Idealized (Weak) Regret Transfer Bound

$$\boxed{\mathrm{regret}^{0\text{-}1}_D [\mathrm{sign} \circ (f - t^*_{D,f})]} \leq \sqrt{2p(1-p) \boxed{\mathrm{regret}^{\mathrm{rank}}_D [f]}}$$

$$\to 0 \qquad\qquad \to 0$$

Bipartite Ranking → Binary Classification

$$f : X \to \mathbb{R}$$

$$t^*_{D,f} \in \operatorname{argmin}_{t \in [-\infty,\infty]} \left\{ \mathrm{er}^{0\text{-}1}_D [\mathrm{sign} \circ (f - t)] \right\}$$

Idealized (Weak) Regret Transfer Bound

$$\mathrm{regret}^{0\text{-}1}_D [\mathrm{sign} \circ (f - t^*_{D,f})] \leq \sqrt{2p(1-p)\, \mathrm{regret}^{\mathrm{rank}}_D [f]}$$

dependence on distribution

Bipartite Ranking → Binary Classification

$$f : X \to \mathbb{R}$$

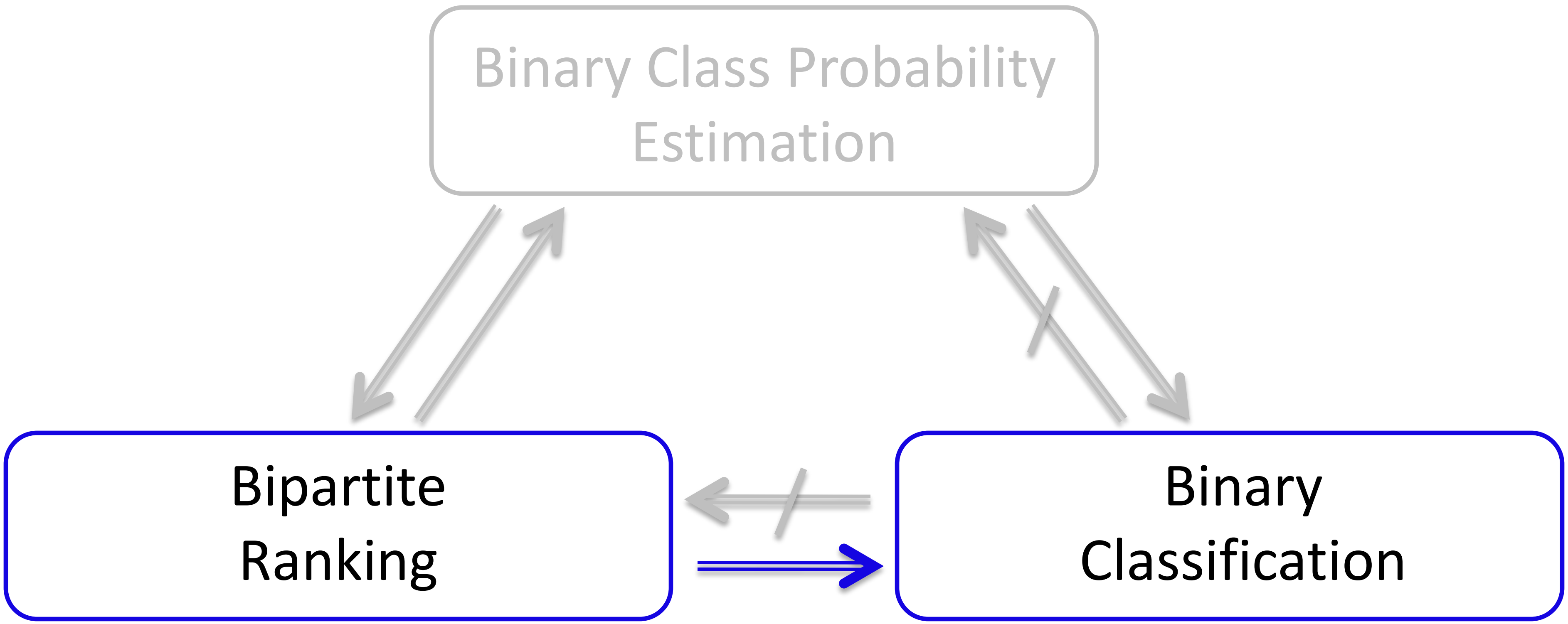
$$t^*_{D,f} \in \operatorname{argmin}_{t \in [-\infty,\infty]} \left\{ \mathrm{er}^{0\text{-}1}_D [\mathrm{sign} \circ (f - t)] \right\}$$

Idealized (Weak) Regret Transfer Bound

$$\mathrm{regret}^{0\text{-}1}_D [\mathrm{sign} \circ (f - t^*_{D,f})] \leq \sqrt{2p(1-p)\, \mathrm{regret}^{\mathrm{rank}}_D [f]}$$

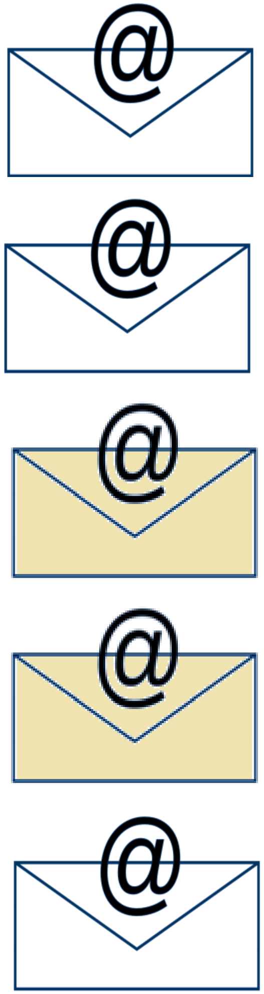
dependence on distribution

Equivalent empirical result

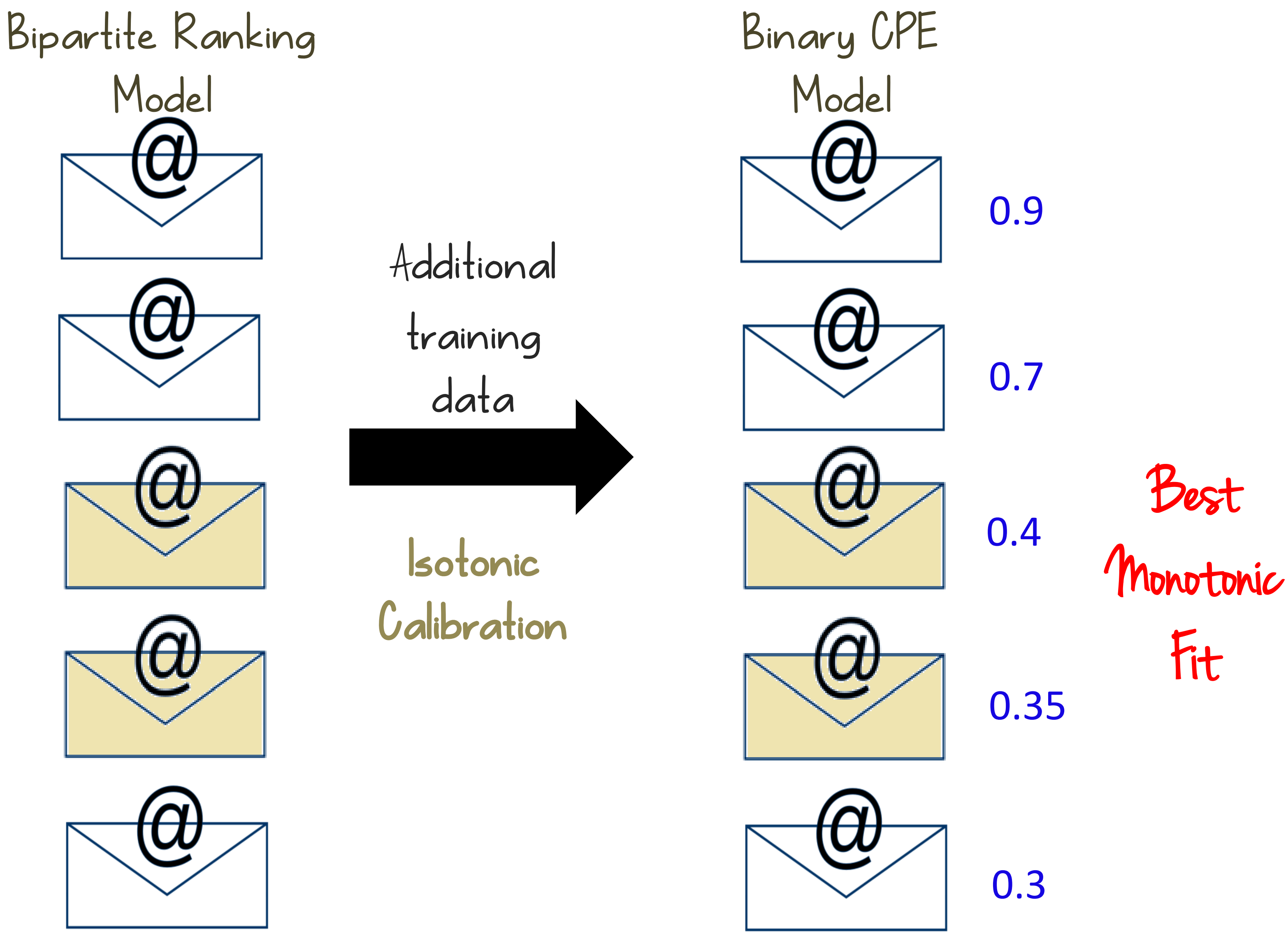

Binary Class Probability Estimation
Bipartite Ranking
Binary Classification

Bipartite Ranking → Binary CPE

Bipartite Ranking Model



Bipartite Ranking → Binary CPE



Bipartite Ranking → Binary CPE

$$f : X \to \mathbb{R}$$

$$\mathrm{Cal}_{D,f} \in \mathrm{argmin}_{g \in \mathcal{G}_{\mathrm{inc}}} \left\{ \mathrm{er}^{\mathrm{sq}}_{D} \left[g \circ f \right] \right\}$$

monotonically increasing functions

Bipartite Ranking → Binary CPE

$$f : X \to \mathbb{R}$$

$$\mathrm{Cal}_{D,f} \in \mathrm{argmin}_{g \in \mathcal{G}_{\mathrm{inc}}} \left\{ \mathrm{er}_D^{\mathrm{sq}}[g \circ f] \right\}$$

monotonically increasing functions

Idealized (Weak) Regret Transfer Bound

$$\mathrm{regret}_D^{\mathrm{sq}}[\mathrm{Cal}_{D,f} \circ f] \ \leq \ \sqrt{8p(1-p)\, \mathrm{regret}_D^{\mathrm{rank}}[f]}$$

Bipartite Ranking $\longrightarrow$ Binary CPE

$$f : X \to \mathbb{R}$$

$$\mathrm{Cal}_{D,f} \in \mathrm{argmin}_{g \in \mathcal{G}_{\mathrm{inc}}} \left\{ \mathrm{er}^{\mathrm{sq}}_D [g \circ f] \right\}$$

monotonically increasing functions

Idealized (Weak) Regret
Transfer Bound

$$\underbrace{\boxed{\mathrm{regret}^{\mathrm{sq}}_D [\mathrm{Cal}_{D,f} \circ f]}}_{\to 0} \leq \sqrt{8p(1-p)\, \underbrace{\boxed{\mathrm{regret}^{\mathrm{rank}}_D [f]}}_{\to 0}}$$

Bipartite Ranking → Binary CPE

$$f : X \to \mathbb{R}$$

$$\mathrm{Cal}_{D,f} \in \mathrm{argmin}_{g \in \mathcal{G}_{\mathrm{inc}}} \left\{ \mathrm{er}^{\mathrm{sq}}_D [g \circ f] \right\}$$

monotonically increasing functions

Idealized (Weak) Regret Transfer Bound

$$\mathrm{regret}^{\mathrm{sq}}_D \left[\mathrm{Cal}_{D,f} \circ f \right] \;\leq\; \sqrt{8p(1-p)\ \mathrm{regret}^{\mathrm{rank}}_D [f]}$$

dependence on distribution

Bipartite Ranking → Binary CPE

$$f : X \to \mathbb{R}$$

$$\mathrm{Cal}_{D,f} \in \mathrm{argmin}_{g \in \mathcal{G}_{\mathrm{inc}}} \left\{ \mathrm{er}_D^{\mathrm{sq}}[g \circ f] \right\}$$

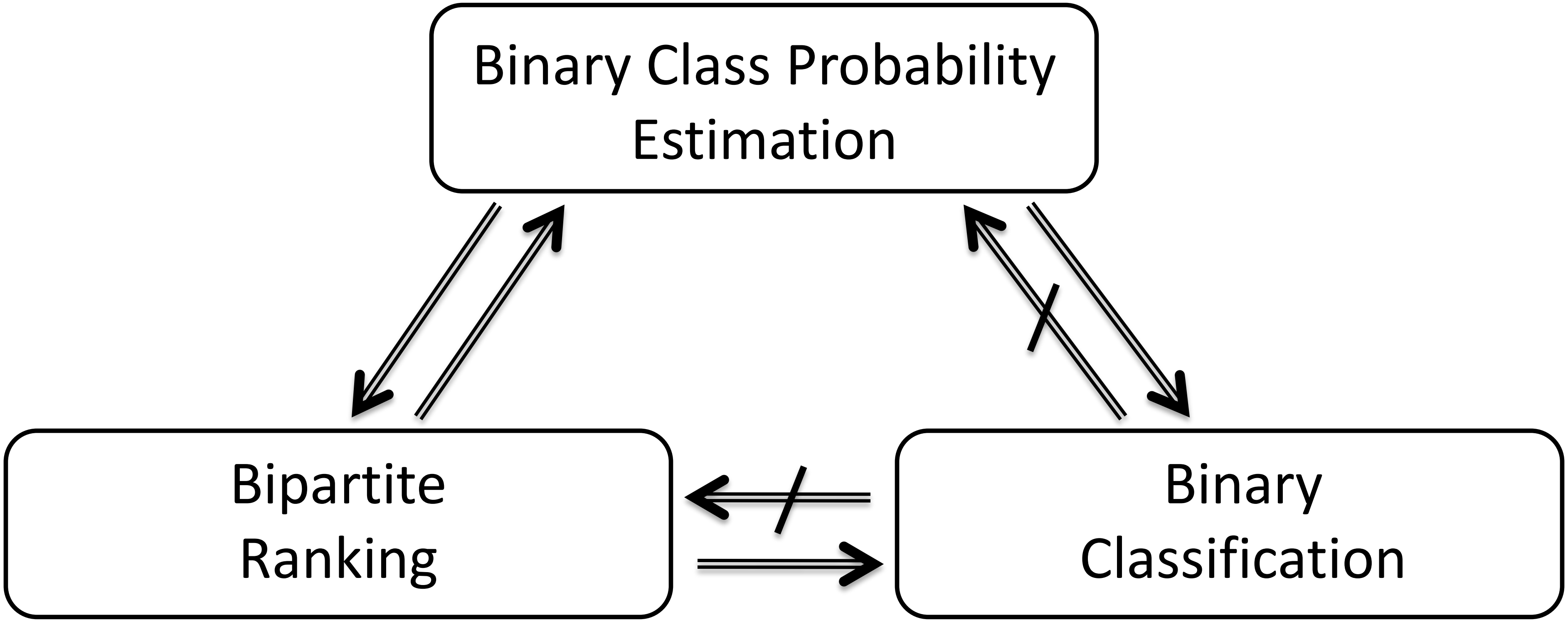
monotonically increasing functions

Idealized (Weak) Regret Transfer Bound

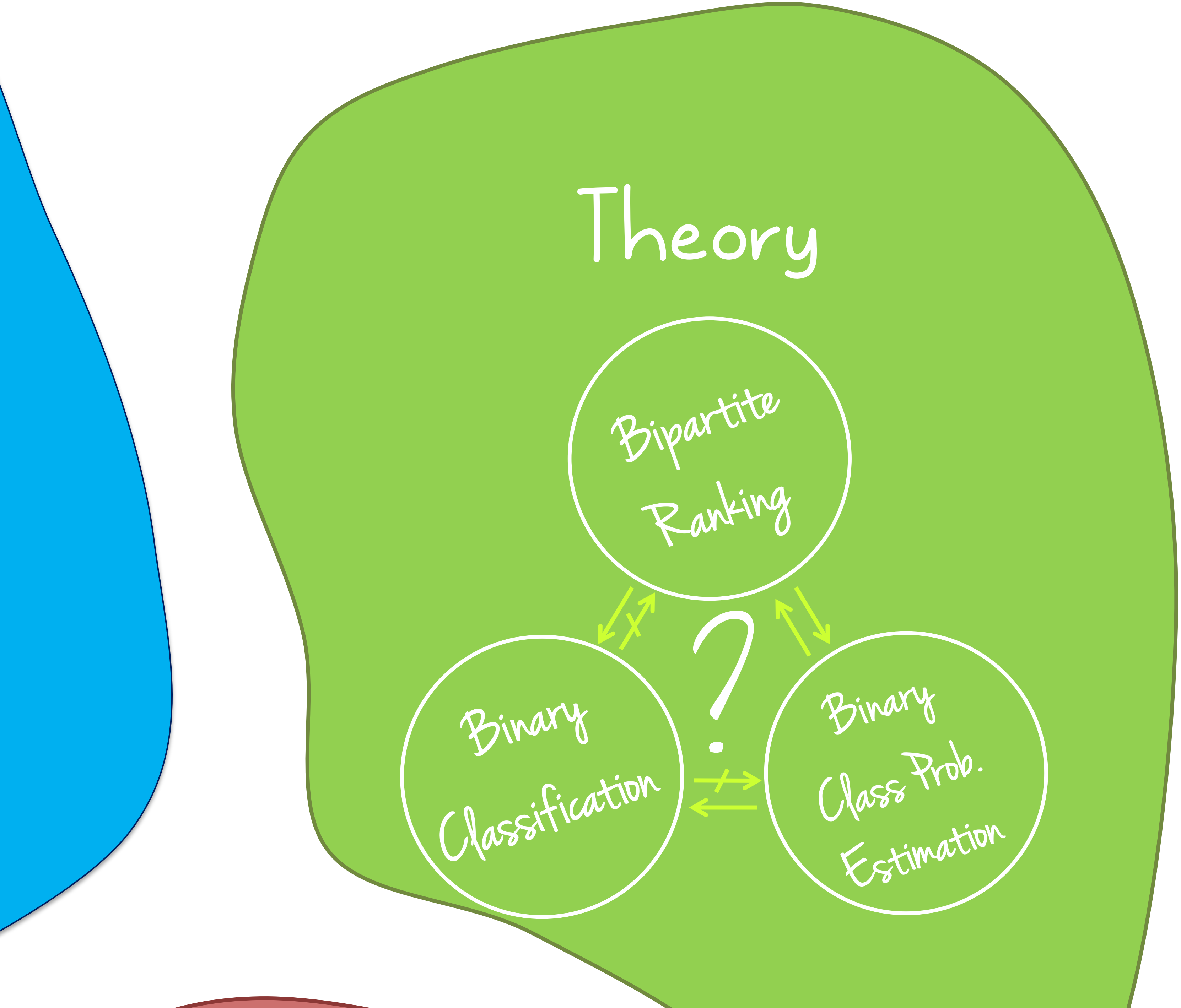
$$\mathrm{regret}_D^{\mathrm{sq}}[\mathrm{Cal}_{D,f} \circ f] \leq \sqrt{8p(1-p)\,\mathrm{regret}_D^{\mathrm{rank}}[f]}$$

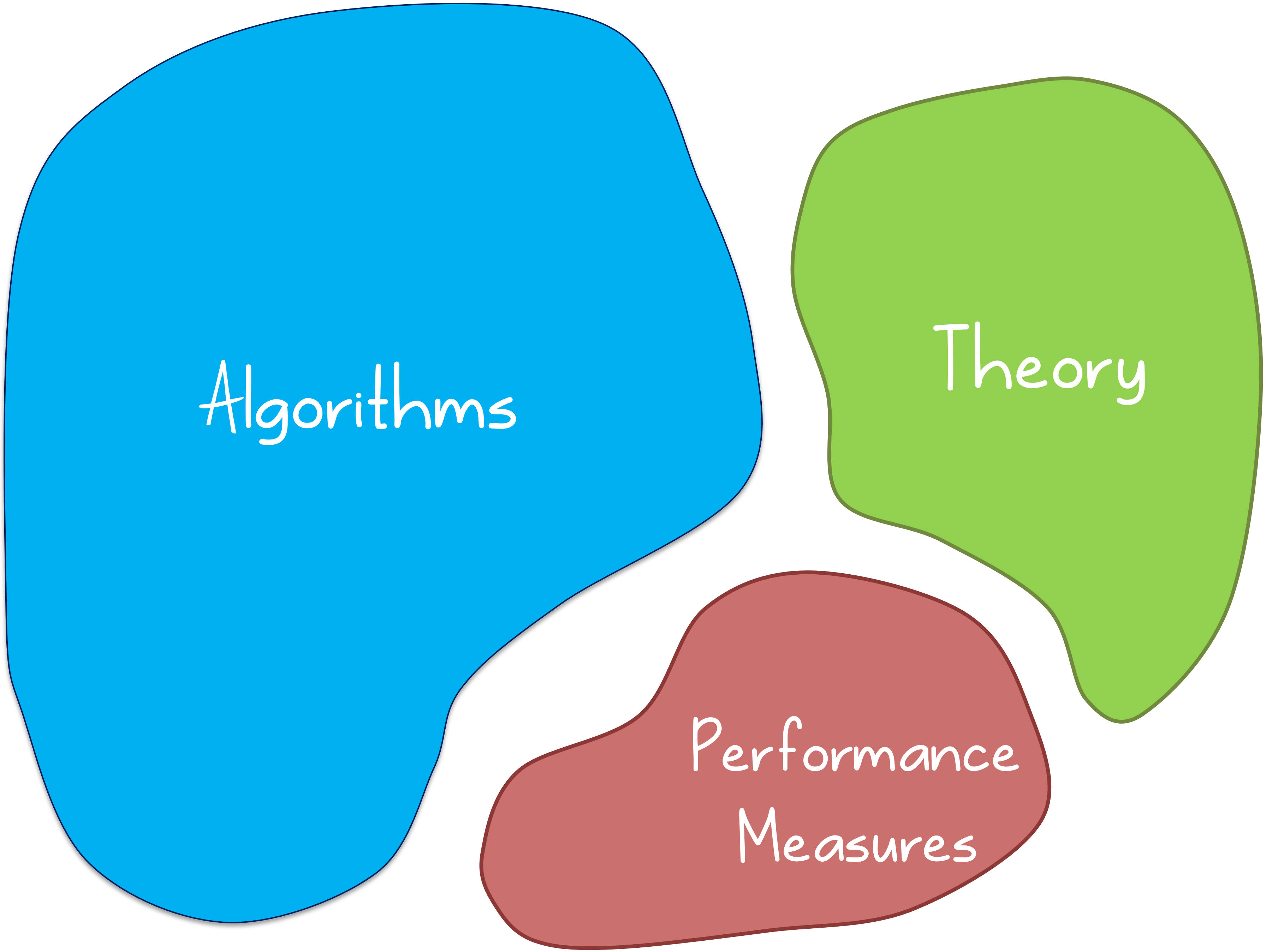
dependence on distribution

Equivalent empirical result

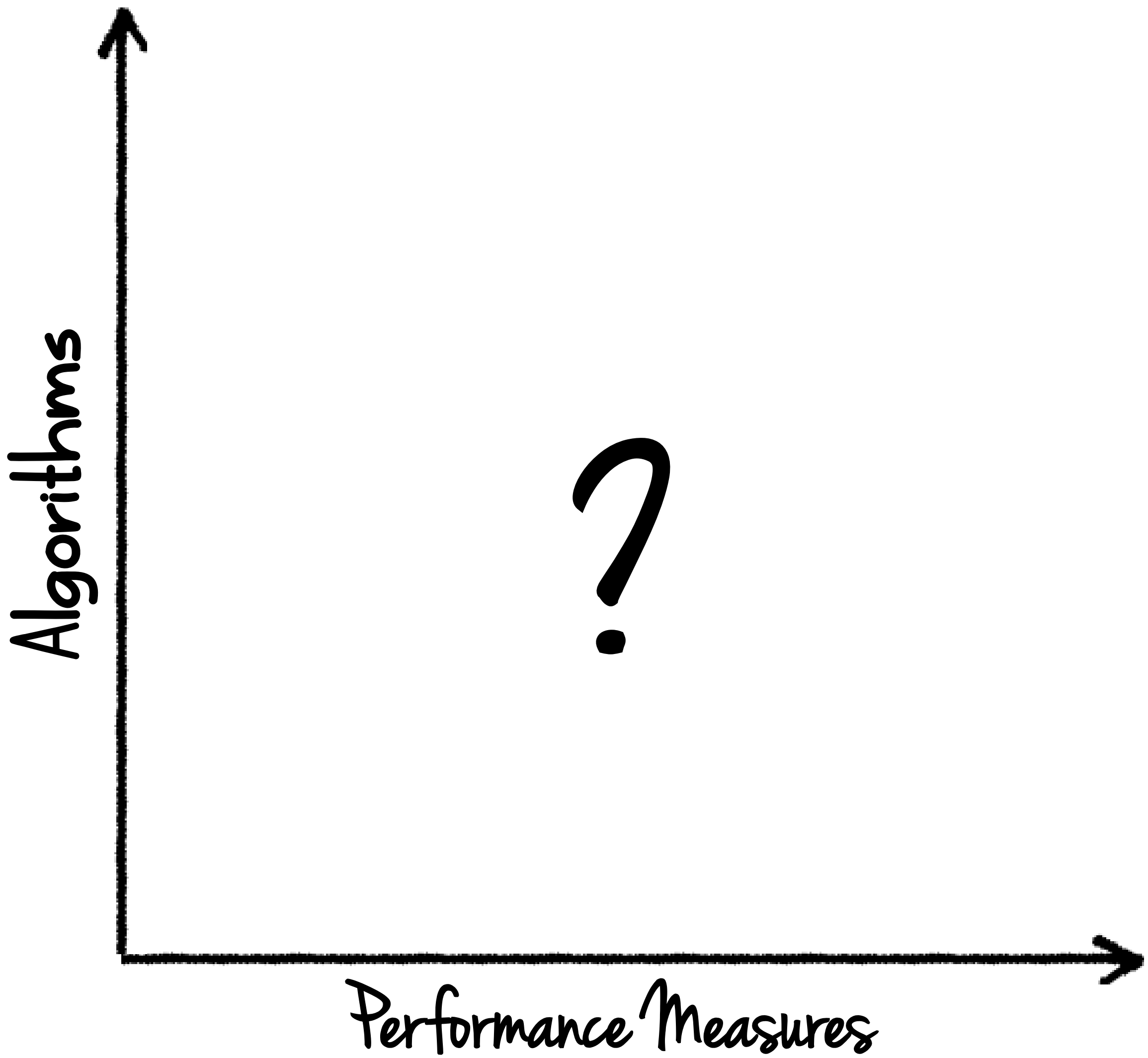


Narasimhan, H. and Agarwal, S. "*On the relationship between binary classification, bipartite ranking, and binary class probability estimation*". In NIPS 2013. To appear.


Theory
Bipartite Ranking
Binary Classification
Binary Class Prob. Estimation
?


Algorithms
Theory
Performance
Measures

take away message ...



open challenges ...

open challenges …

- Designing statistical consistent algorithms for multivariate performance measures

open challenges …

- Designing statistical consistent algorithms for multivariate performance measures

- Online algorithms for optimizing multivariate performance measures

open challenges ...

- Designing statistical consistent algorithms for multivariate performance measures
- Online algorithms for optimizing multivariate performance measures
- Structural SVM approach:
  - Conditions for statistical consistency
  - Understanding upper bound optimized

Acknowledgements

Machine Learning and Learning Theory Group @ IISc

Shivani Agarwal

Harish Guruprasad Ramasamy

Siddarth Ramamohan

Arun Rajkumar

Rohit Vaish

Arpit Agarwal

Saneem Ahmed

Suprovat Ghoshal

Aadirupa Saha

Acknowledgements

Collaborators

Sanjay Chawla, University of Sydney and NICTA, Sydney

Pradip k. Majumder, Mitra Biotech, Bangalore

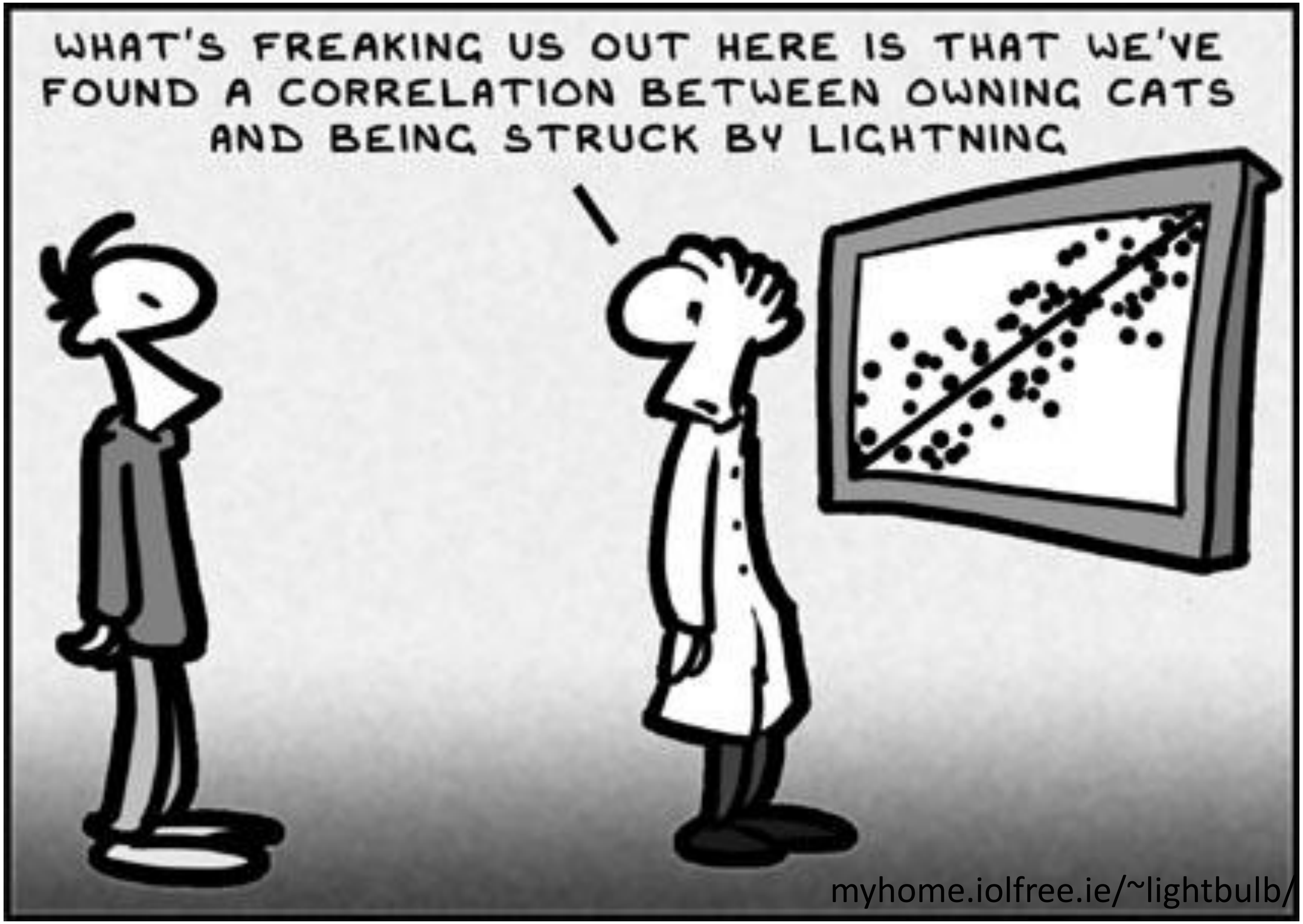
Biswanath Majumder, Mitra Biotech, Bangalore

Aditya Menon, NICTA, Sydney

Padhma Radhakrishnan, Mitra Biotech, Bangalore

Shiladitya Sengupta, Harvard Medical School, Boston

Mallikarjun Sundaram, Mitra Biotech, Bangalore


WHAT'S FREAKING US OUT HERE IS THAT WE'VE FOUND A CORRELATION BETWEEN OWNING CATS AND BEING STRUCK BY LIGHTNING
myhome.iolfree.ie/~lightbulb/

Questions?

Most images in this presentation were obtained using
Google image search (http://www.google.co.in/imghp)

Some of the hand-drawn shapes & clipart images were obtained from the following sites:

http://www.clker.com/

http:// www.free-power-point-templates.com /

http://neilslorance.wordpress.com/

Some fonts in this presentation were downloaded from:

http://www.dafont.com/