

On the Relationship Between Binary Classification, Bipartite Ranking, and Binary Class Probability Estimation

Harikrishna Narasimhan

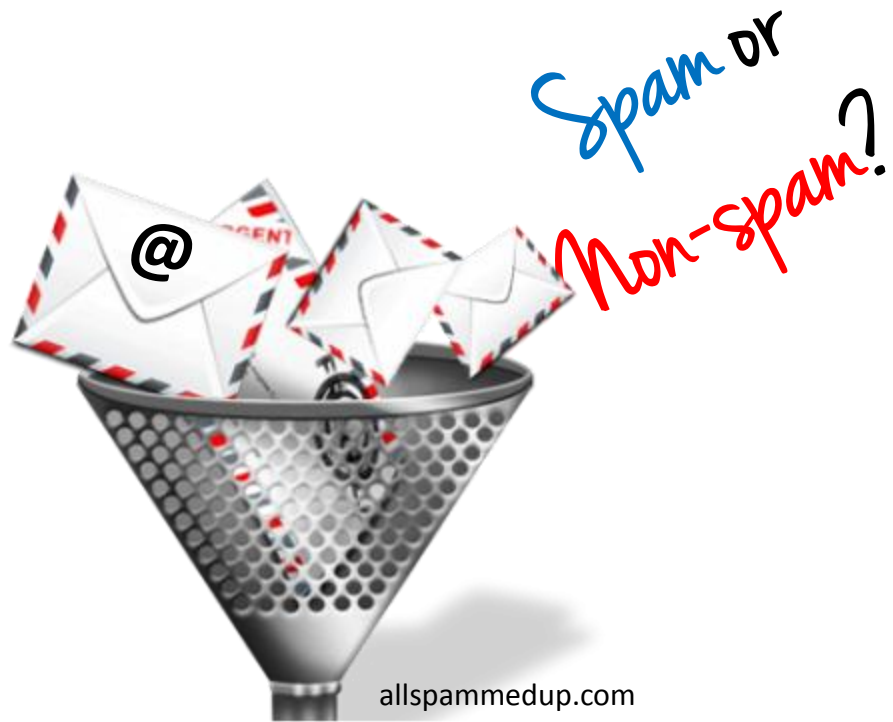


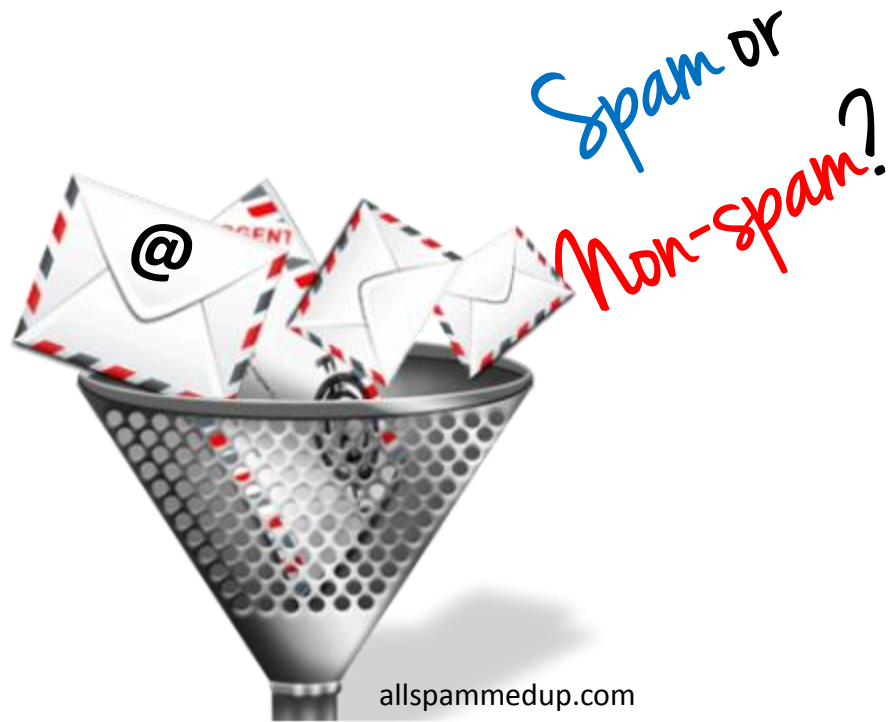
Department of Computer Science and Automation
Indian Institute of Science, Bangalore

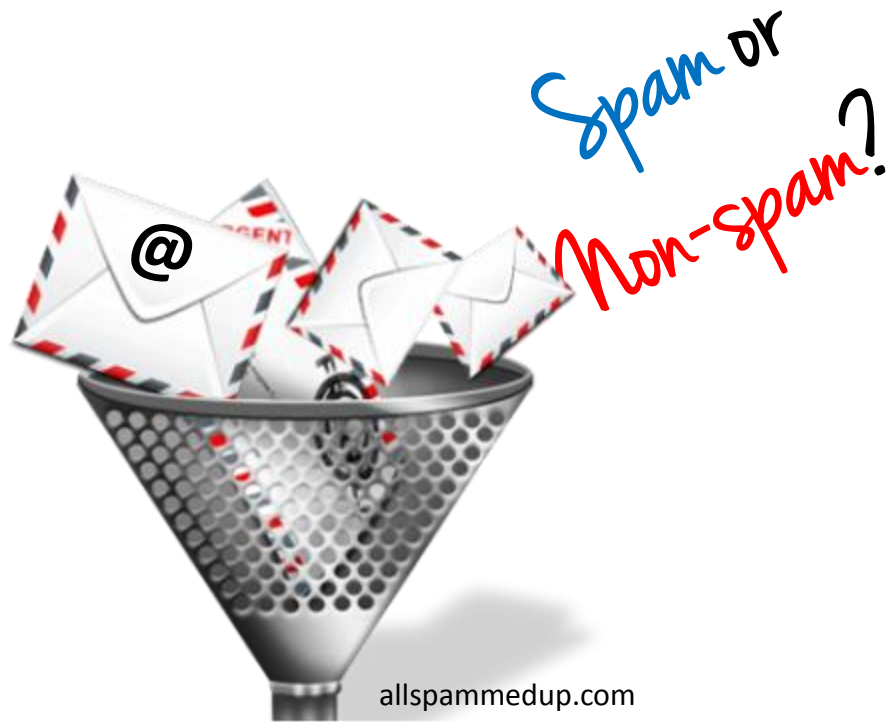
Spam or
Non-spam?



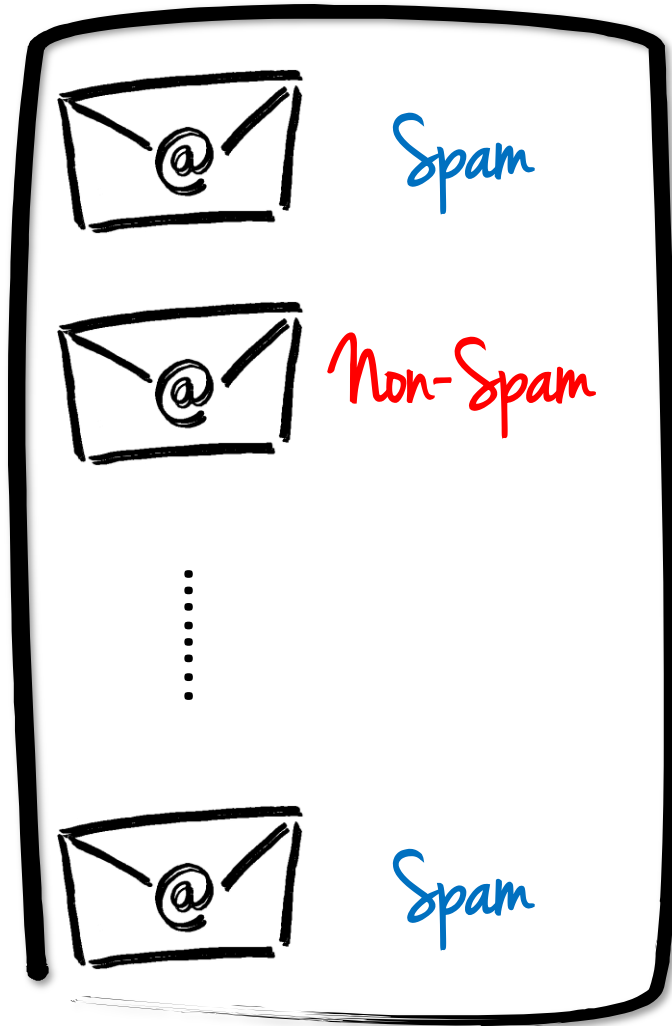
allspammedup.com





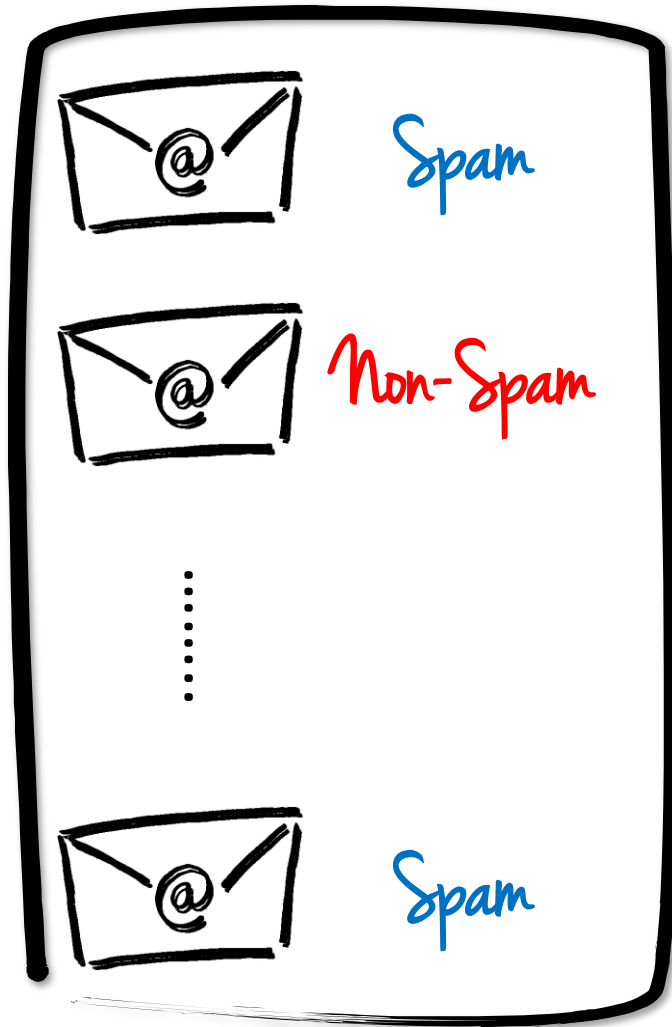


Learning from 'Binary' Labels



Training Set

Learning from 'Binary' Labels



Training Set



Binary Classification

Google

in:spam



Gmail ▾



Delete forever

Not spam



More ▾

COMPOSE

Time travelers PLEASE HELP!!!!!!

Spam x



Inbox (12)

Chats

Sent Mail

Drafts (229)

Spam (176)

▸ Circles

Personal

Project K

Work

More ▾

<xxxxxxx@yyy.com>

10:29 PM (1 minute ago) ☆

to me ▾

If you are a time traveler or alien disguised as human and or have the technology to travel physically through time I need your help!

My life has been severely tampered with and cursed!!

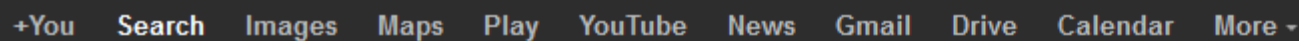
I have suffered tremendously and am now dying!

I need to be able to:

Travel back in time.

Spam or Non-spam?

Bipartite Ranking



learning to rank



Web

Images

Maps

More ▼

Search tools

About 142,000,000 results (0.32 seconds)

[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

[PDF] Future directions in **learning to rank** - Journal of Machine Learning ...

jmlr.org/proceedings/papers/v14/chapelle11b/chapelle11b.pdf ▼

by O Chapelle - Cited by 26 - Related articles

Abstract. The results of the **learning to rank** challenge showed that the quality of the ... can be made in the “core” and traditional problematics of **learning to rank**.

[PDF] **Learning to Rank** for Information Retrieval This Tutorial - WWW 2009

www2009.org/.../T7A-LEARNING%20TO%20RANK%20TUTORIAL.p... ▼

Apr 12, 2009 - ... other conferences and journals might not be covered comprehensively.
4/20/2009. Tie-Yan Liu @ WWW 2009 Tutorial on **Learning to Rank**.

[PDF] **Ranking Methods in Machine Learning** - Computer Science and ...

www.csa.iisc.ernet.in/~shivani/Events/SDM-10.../sdm10-tutorial.pdf ▼

Ranking Methods in Machine Learning. A Tutorial Introduction. Shivani Agarwal. Computer Science & Artificial Intelligence Laboratory, Massachusetts Institute of ...

Relevant documents

above

Irrelevant documents

<http://www.google.com/>

Binary Class Probability Estimation

New User? Register | Sign In | Help

Get new Yahoo! Mail apps

MY YAHOO!

Web Images News

Web Search

My Main Tab New Tab

Content Themes Options

Sign in to permanently save changes you make on My Yahoo!

Sign In I'm new here.

Weather 24°C Horoscope Stocks Bookmarks

Quicklinks

Yahoo! Services

Answers Astrology Buzz Cricket Dalnig Jagran Finance Filokir Games Groups Lifestyle Local Mail Maps Messenger Mobile Movies News Pulse Weather

More Yahoo! Services

Today

TODAY - August 26, 2013

Sachin retires
15,000-runs summit
Tendulkar: 100 Not Out

Sachin's tall claim John crashes superbike Mumbai's horror story Live: Yahoo! Real Time Fashion faux pas pros One night in Bangkok Chennai's 'fire baby' Rape: How the case unfolded

1 of 48

National News | National News Headlines - Yahoo! India News

- BJP wants FIR against Dikshit, ministers in transport 'scam' - 36 minutes ago
- Bengal assembly passes bill for time-bound services - 30 minutes ago
- Judiciary, Politicians, Police responsible for dismal rates of rape conviction: Brinda Karat - 37 minutes ago
- Mumbai police to seek 'maximum punishment' for rape case accused - 37 minutes ago
- VHP yatra not a political agenda, saints have right to go to Ayodhya: Singhal - 37 minutes ago

Bollywood News & Gossips | Regional Movie News Headlines - Yahoo! Movies India

Options

Options

Cricket Scores and Schedules

Live Now Results Upcoming

Aug 24 ZIM vs. PAK
2nd T20I - Harare - View Scores
PAK lead the match by 49 runs
ZIM 160/6 (20) - 2nd Inn

What is the probability of a user clicking this ad?

Get a FREE Equity Trading Account at FundsIndia.com

No Equity Account Opening FEE of ~~Rs 500~~ Till August 31st, 2013*

Name Email
***** Mobile

REGISTER NOW

Disclaimer: FundsIndia collects the information within this form solely responsible for their use of it. Please read their Privacy Policy. FundsIndia does not collect or use this information in any way.

*Mutual Fund Investment

http://in.yahoo.com

three fundamental problems ...

Binary
Classification

Bipartite
Ranking

Binary
Class Prob.
Estimation

three fundamental problems ...

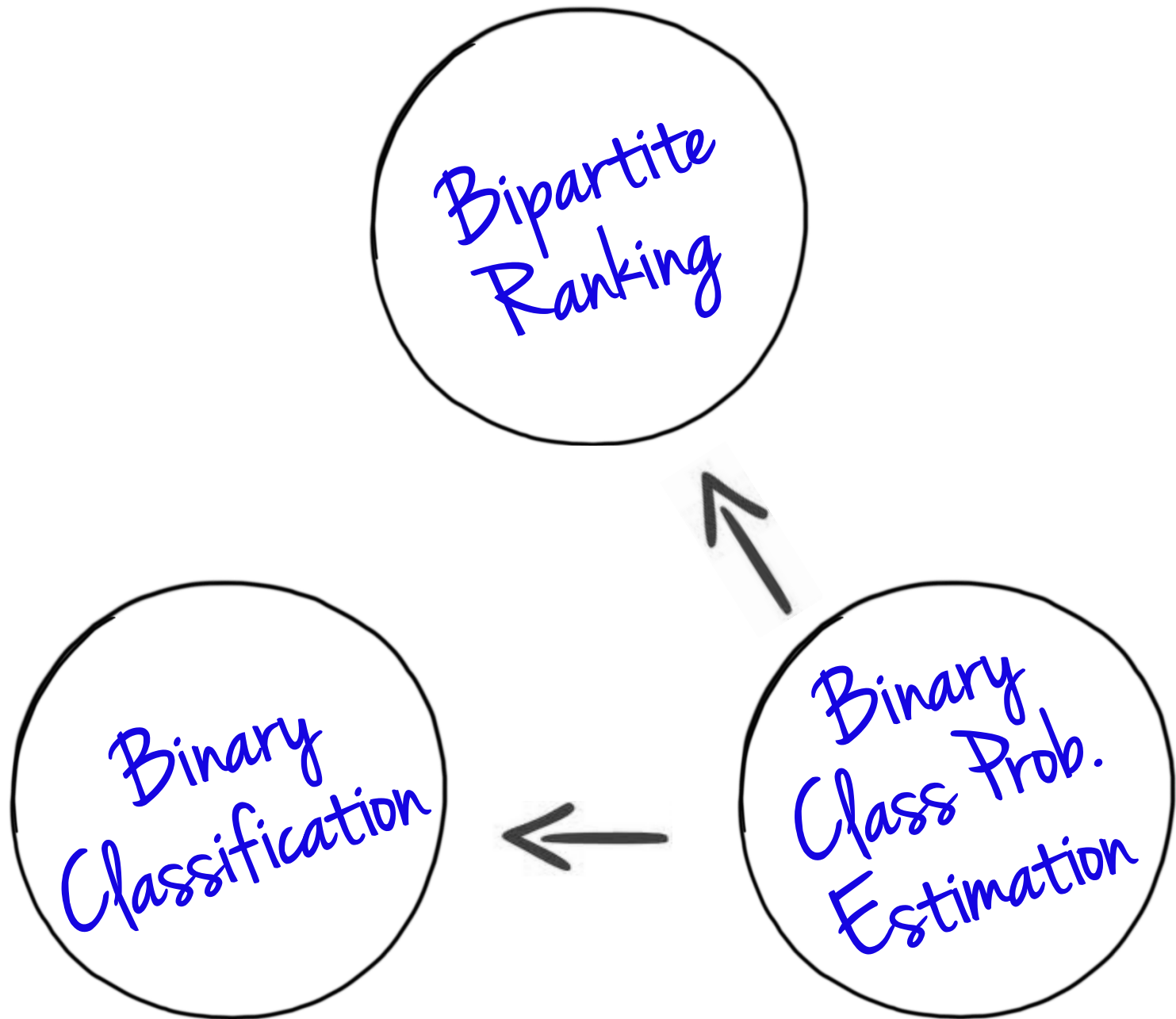
Binary
Classification

Bipartite
Ranking

Binary
Class Prob.
Estimation

Statistical Consistency?

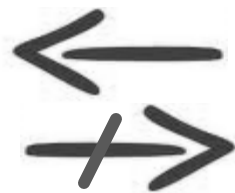


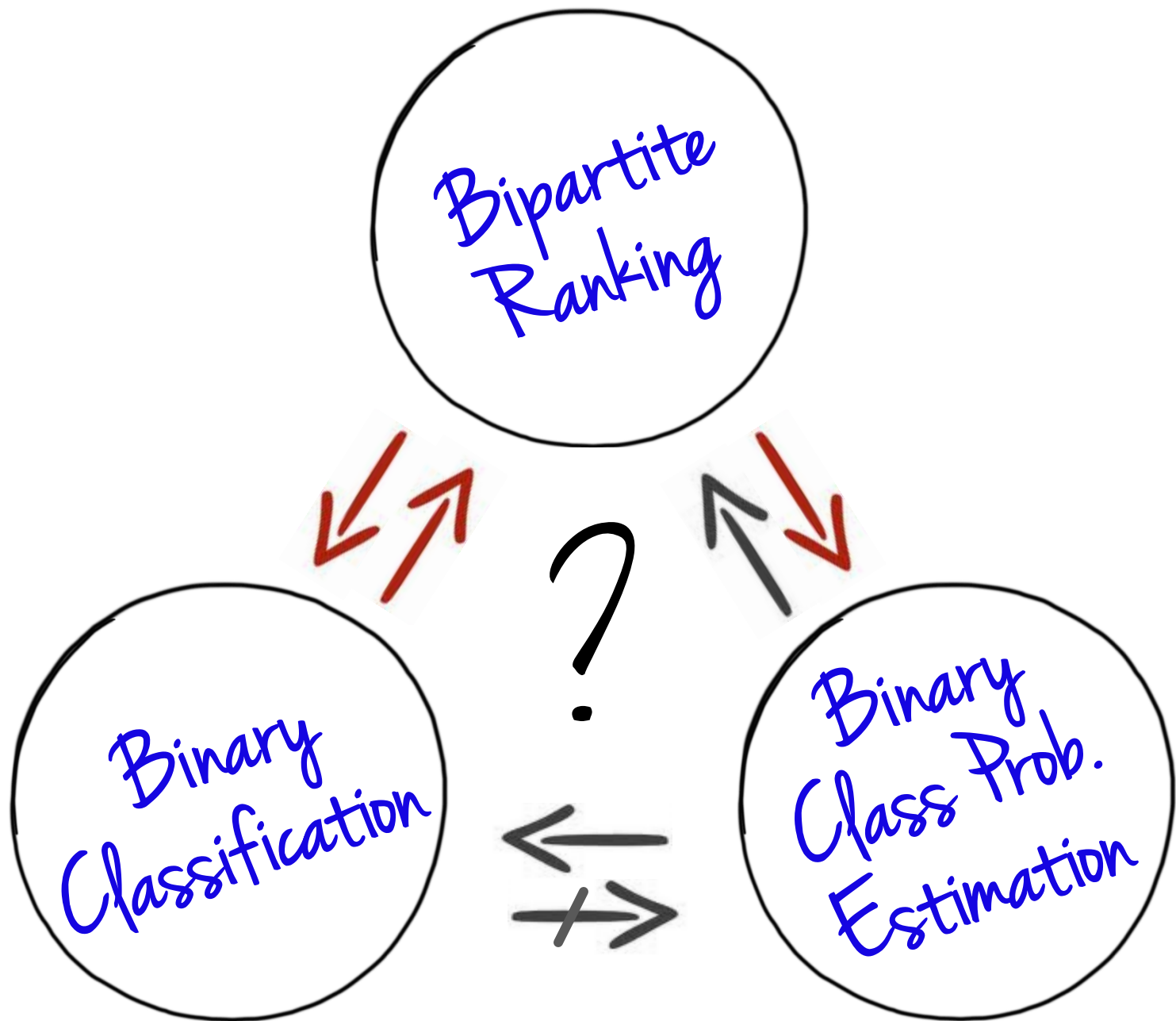


Binary
Classification

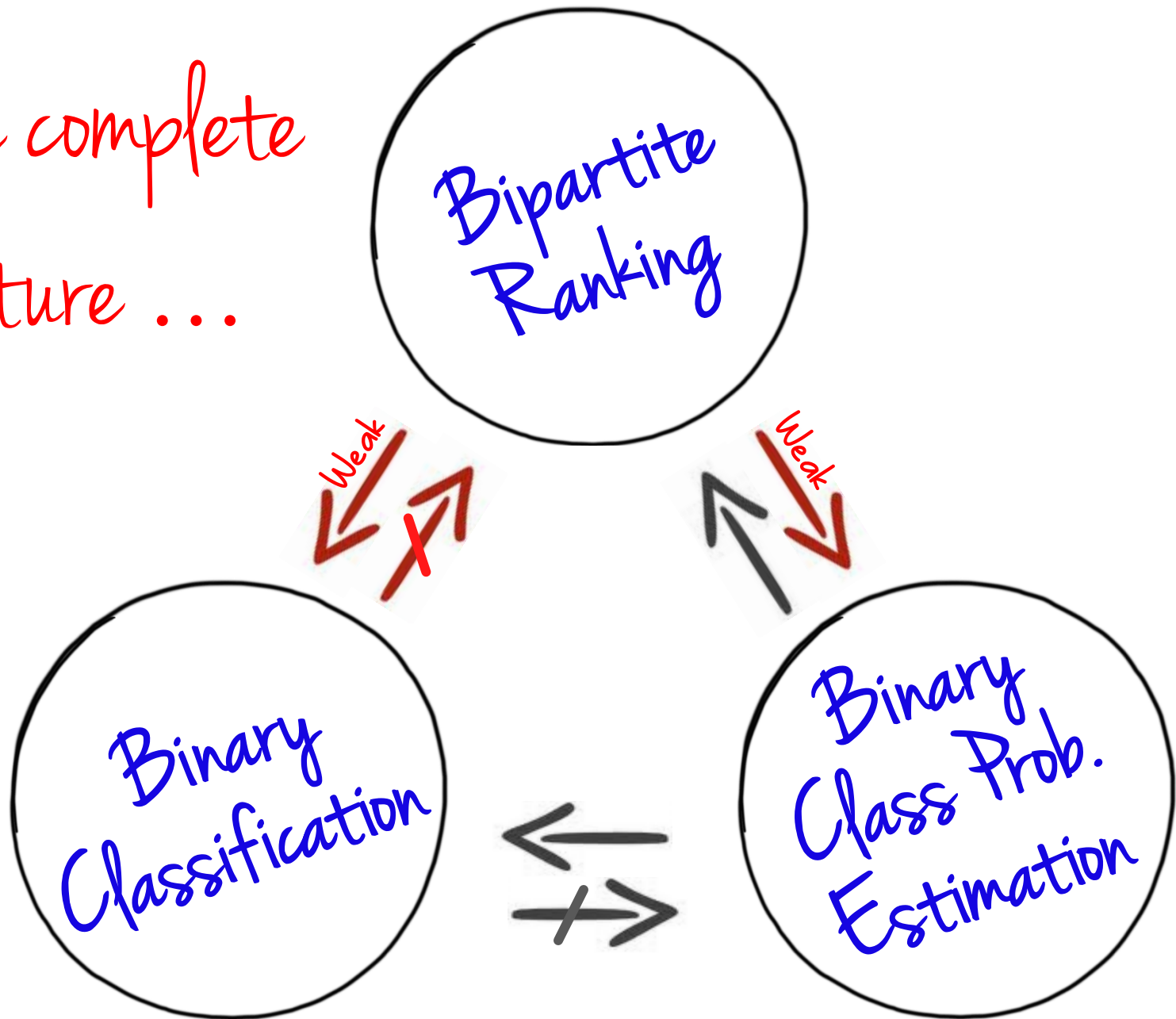
Bipartite
Ranking

Binary
Class Prob.
Estimation

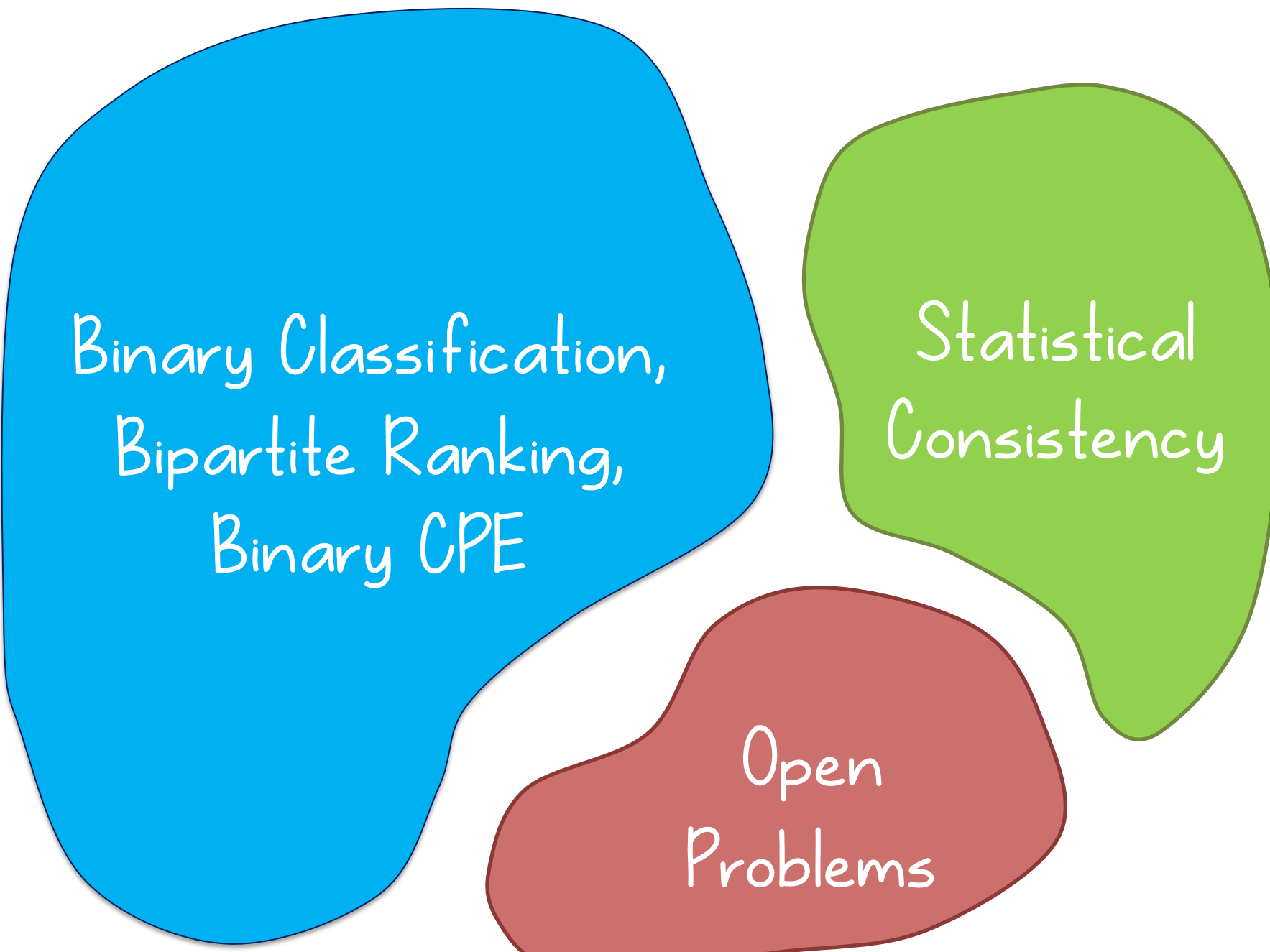




the complete
picture ...



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,
Bipartite Ranking,
Binary CPE

Statistical
Consistency

Open
Problems

Statistical Consistency

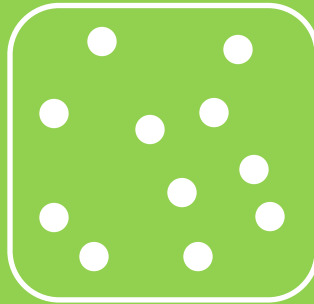
Binary
Classification

Bipartite
Ranking

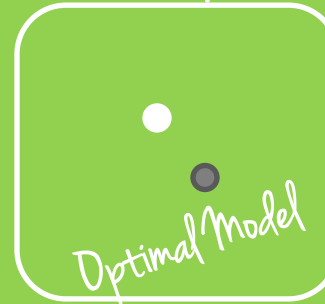
Binary
Class Prob.
Estimation

Statistical Consistency

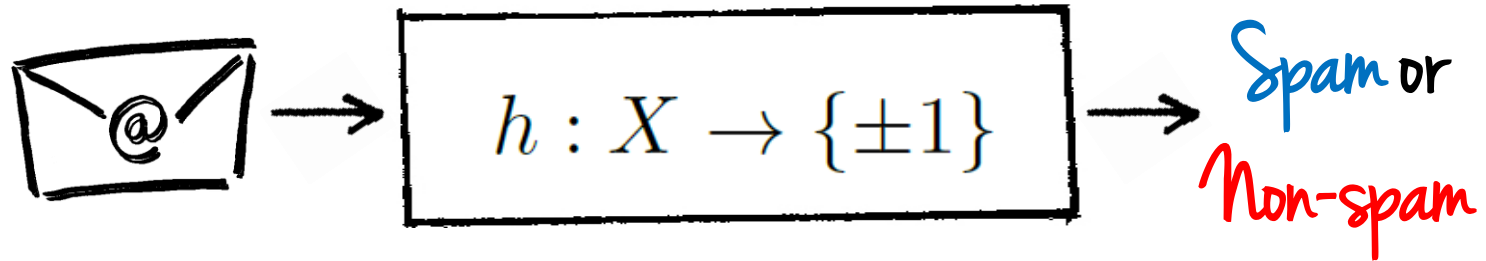
Data Space



Model Space



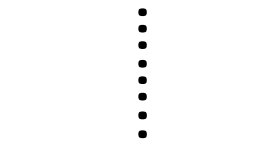
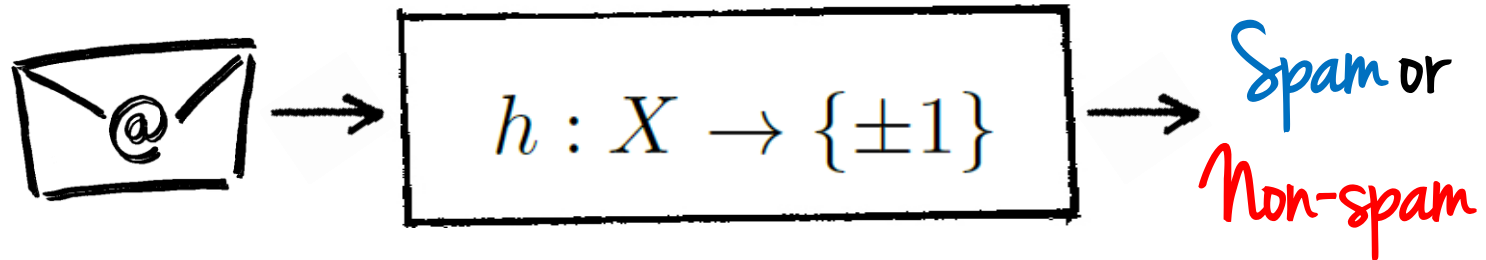
Binary Classification



...

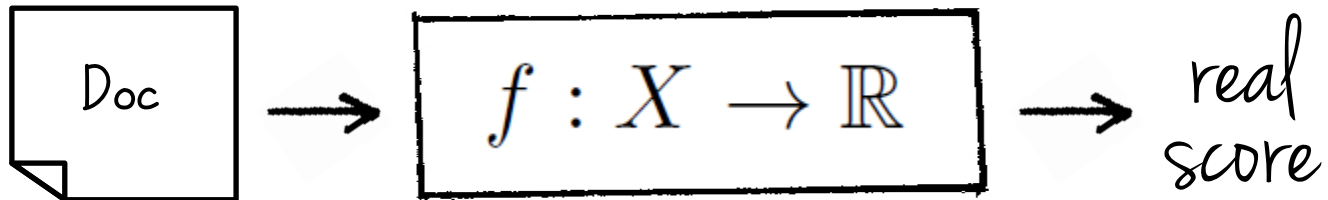


Binary Classification



0-1
Classification
Error

Bipartite Ranking



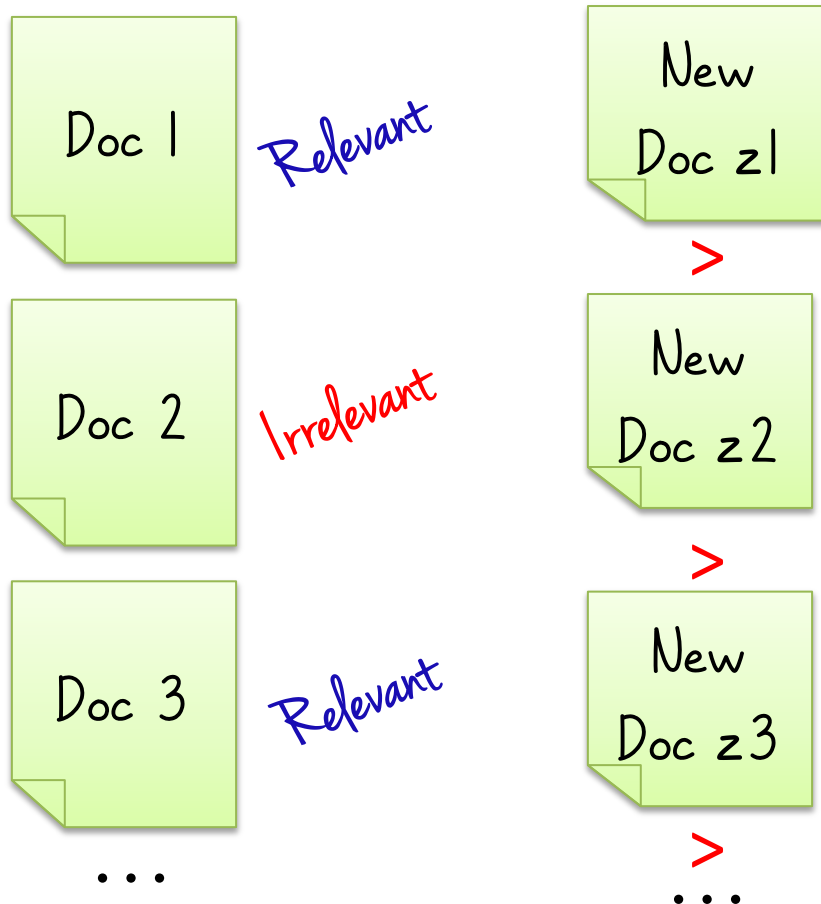
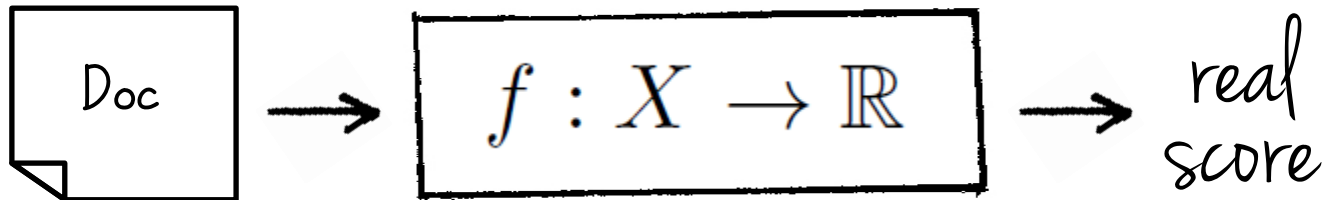
Doc 1 *Relevant*

Doc 2 *Irrelevant*

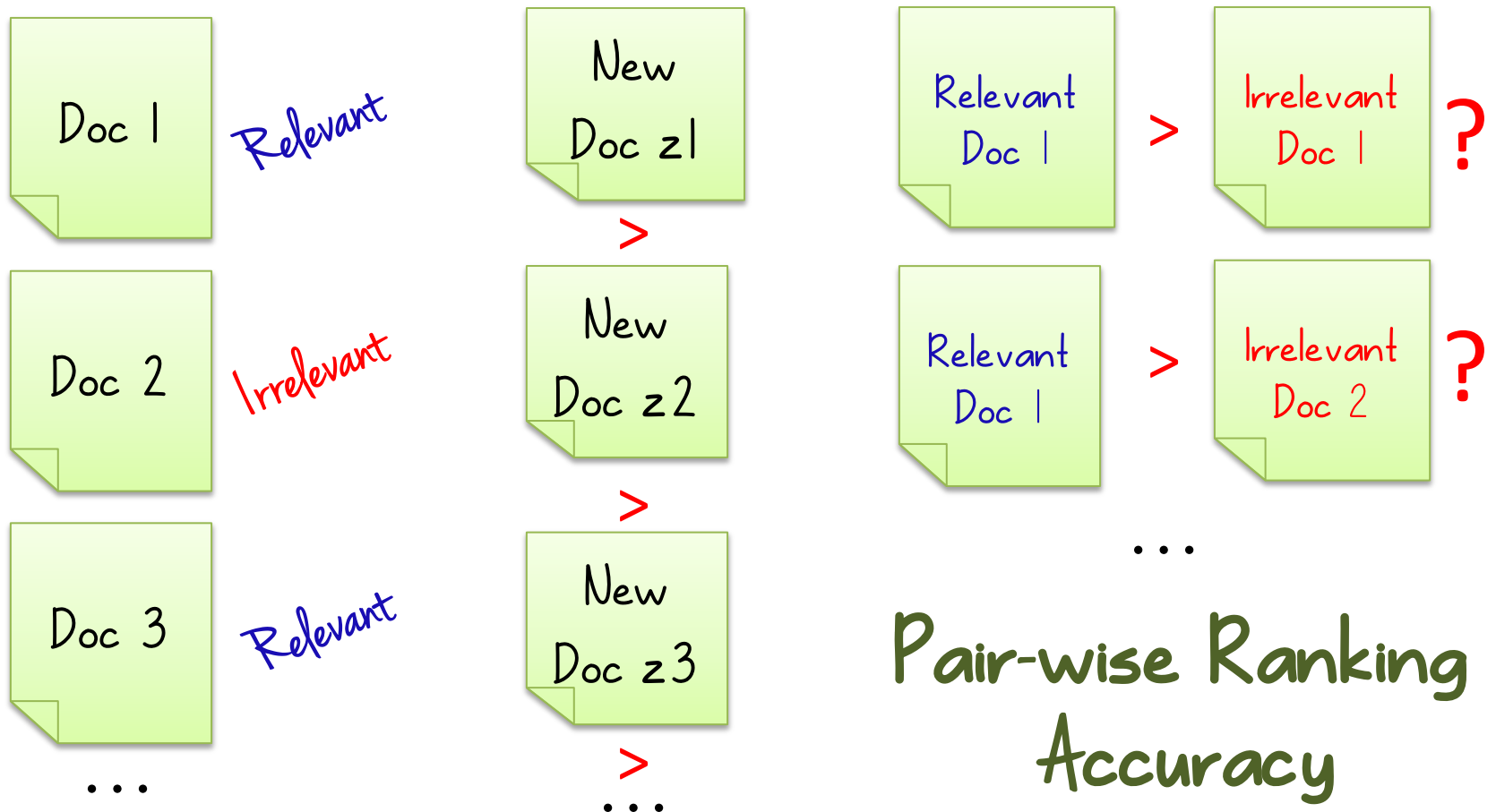
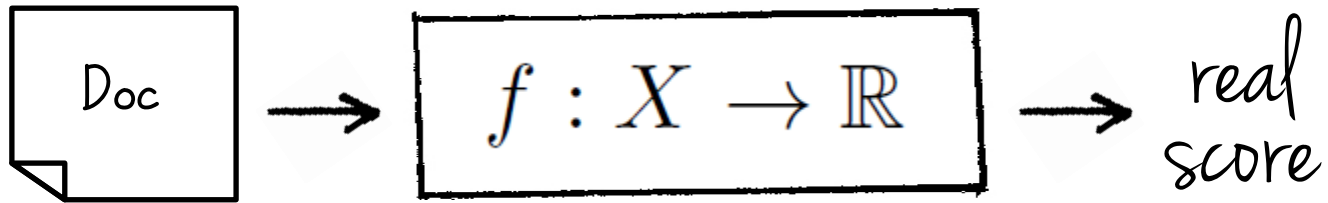
Doc 3 *Relevant*

...

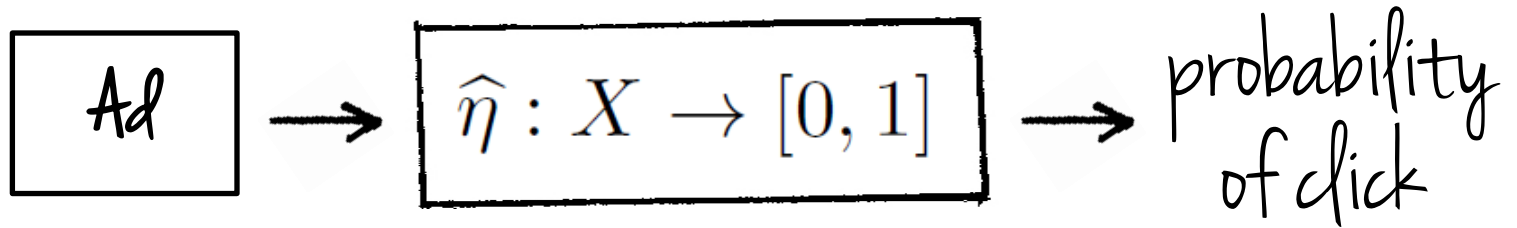
Bipartite Ranking



Bipartite Ranking



Binary Class Probability Estimation



Ad 1

Click

Ad 2

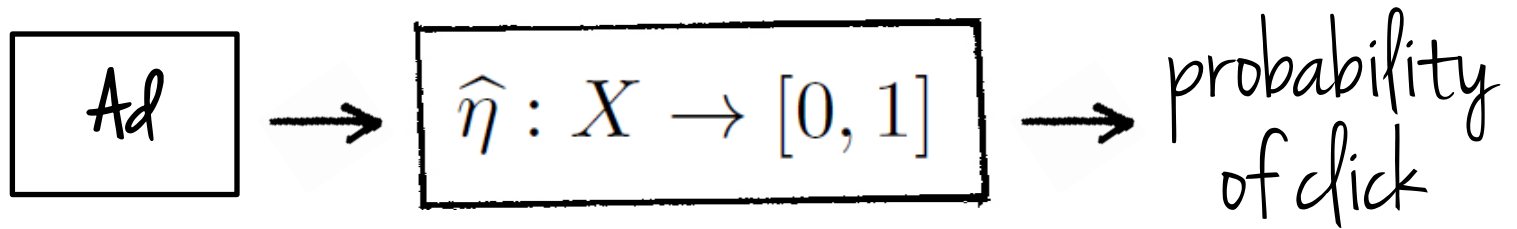
No Click

Ad 3

Click

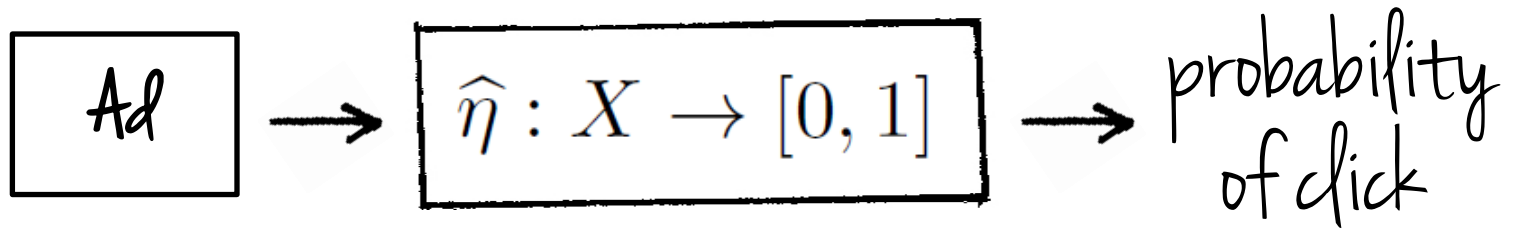
...

Binary Class Probability Estimation



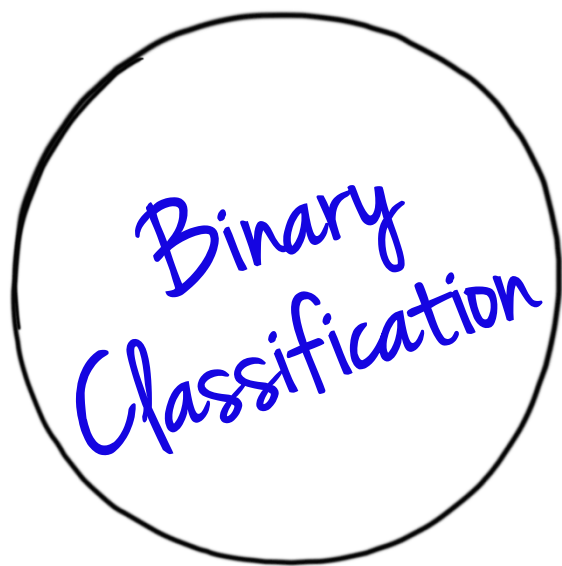
Ad 1	Click	New Ad 1	0.1
Ad 2	No Click	New Ad 2	0.4
Ad 3	Click	New Ad 3	0.2
...		...	

Binary Class Probability Estimation



Ad 1	Click	New Ad 1	0.1	0.2
Ad 2	No Click	New Ad 2	0.4	0.45
Ad 3	Click	New Ad 3	0.2	0.1
...		...		

Squared Error



0-1
Classification
Error



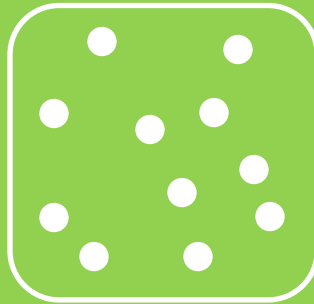
Pair-wise
Accuracy



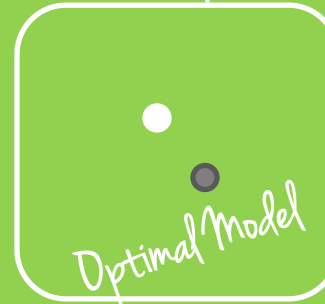
Squared
Error

Statistical Consistency

Data Space

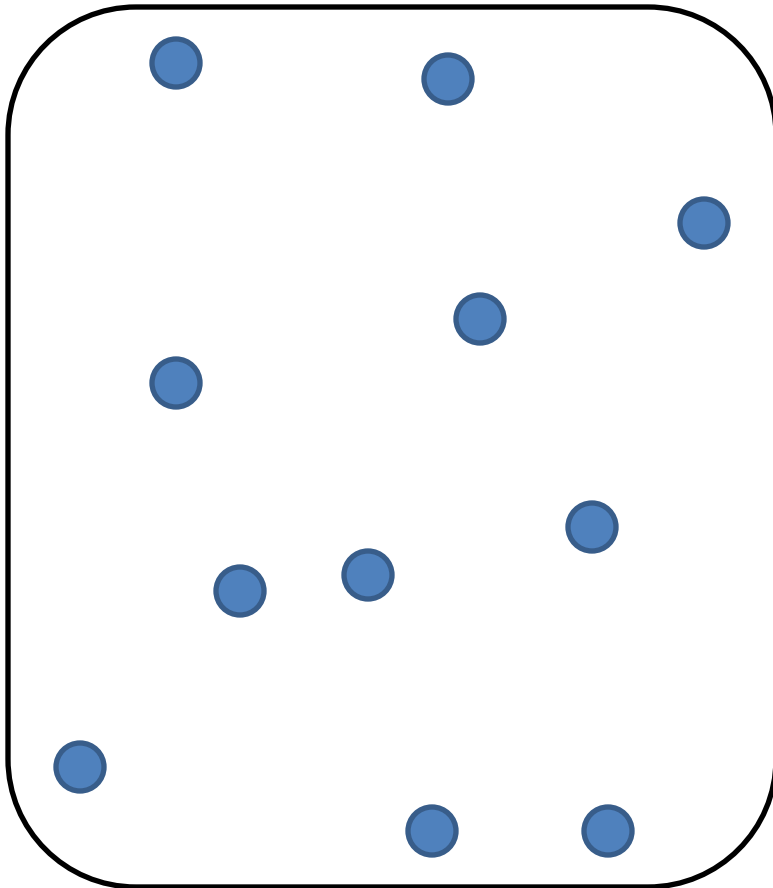


Model Space

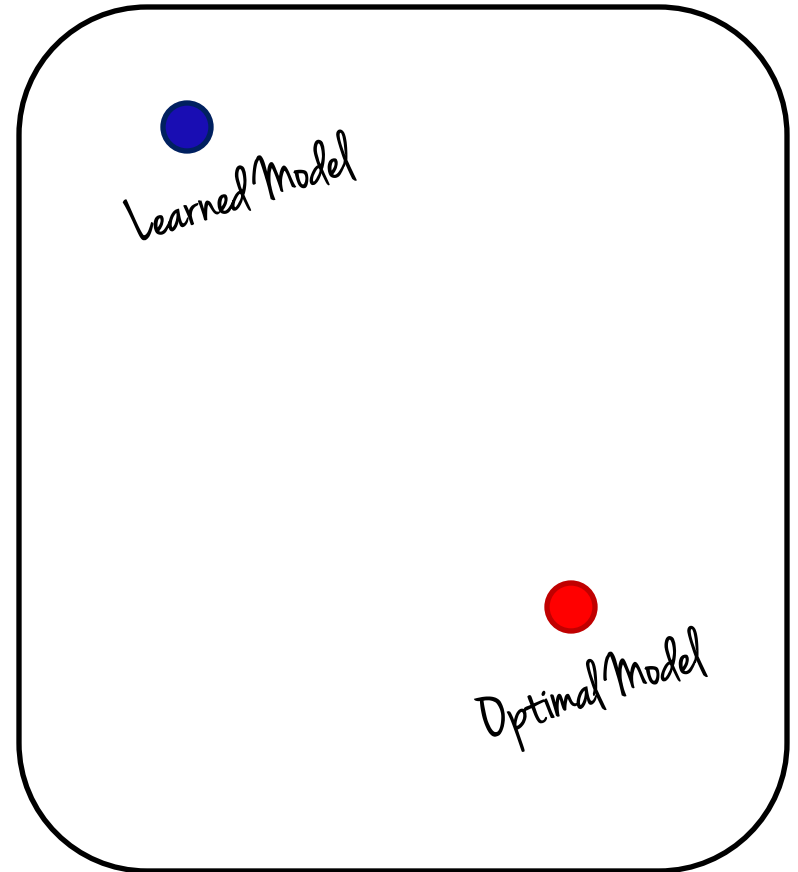


Statistical Consistency

Data Space

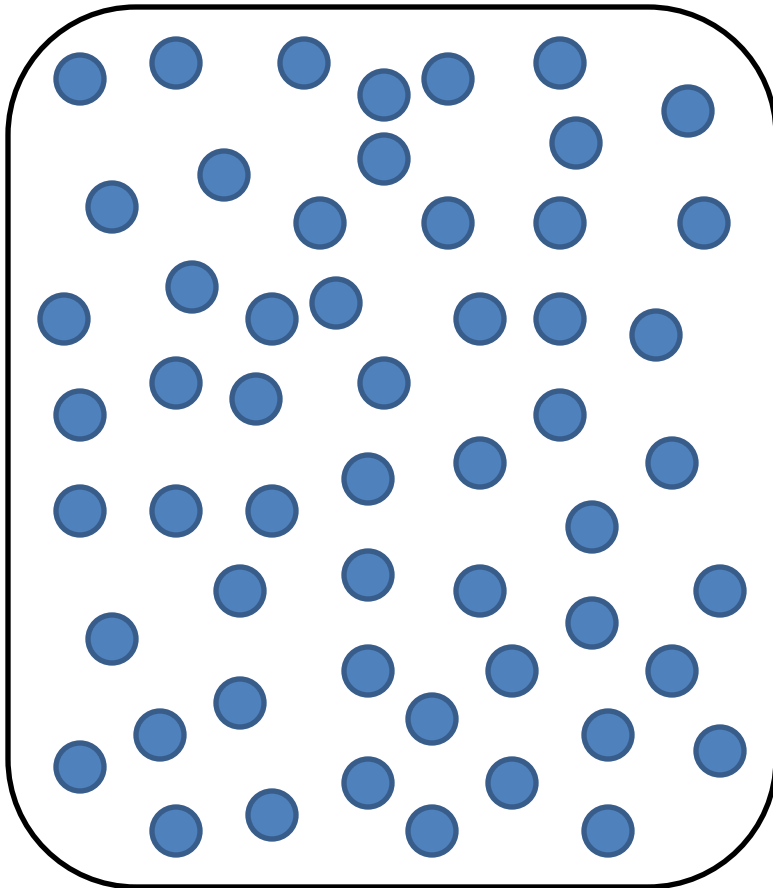


Model Space

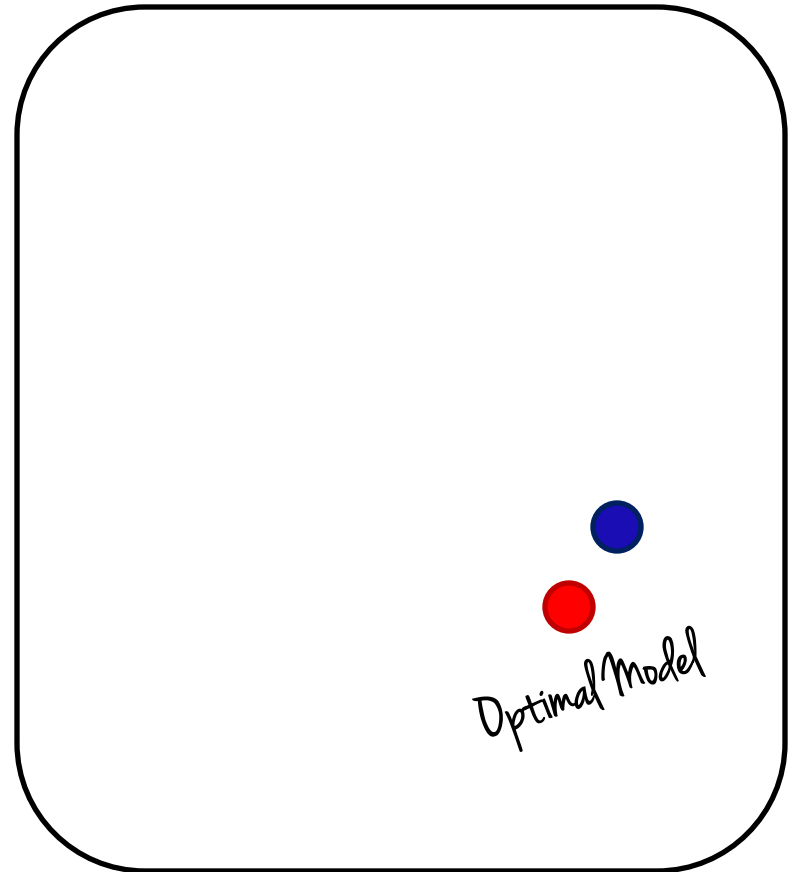


Statistical Consistency

Data Space

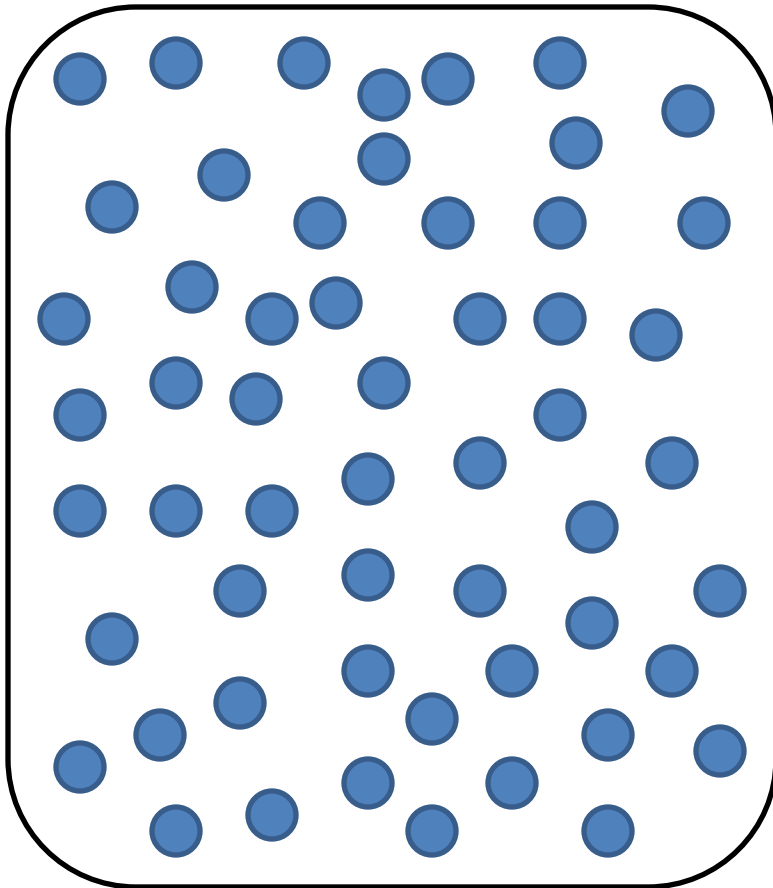


Model Space

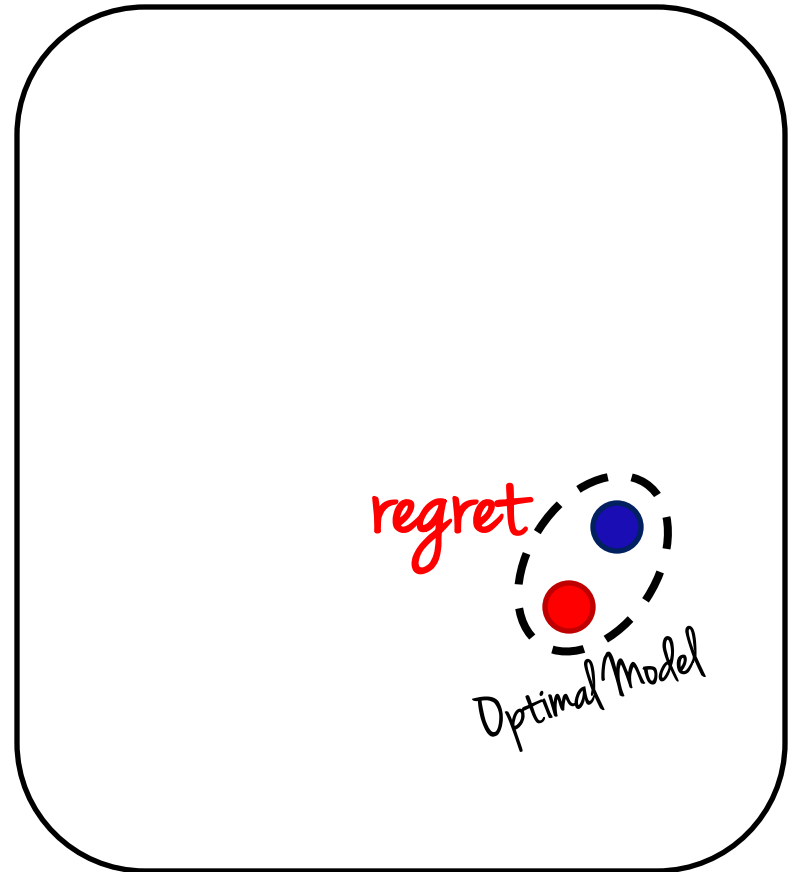


Statistical Consistency

Data Space



Model Space



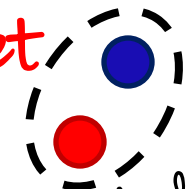
Statistical Consistency

Data Space

Model Space

regret $\xrightarrow{P} 0$?

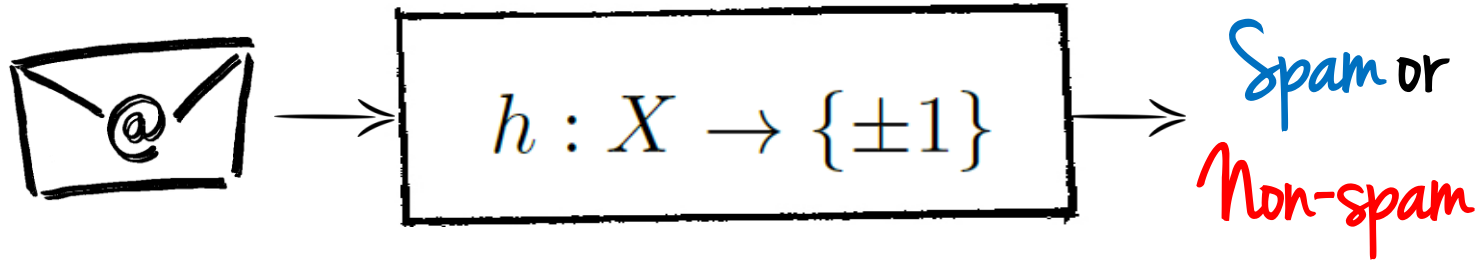
regret



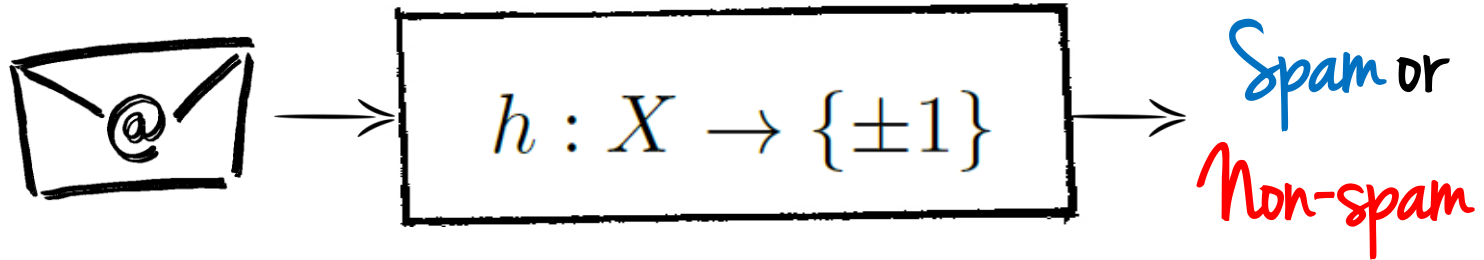
Optimal Model

The diagram shows a dashed circle containing a red dot and a blue dot. The red dot is labeled 'Optimal Model' and the blue dot is labeled 'regret'.

Binary Classification

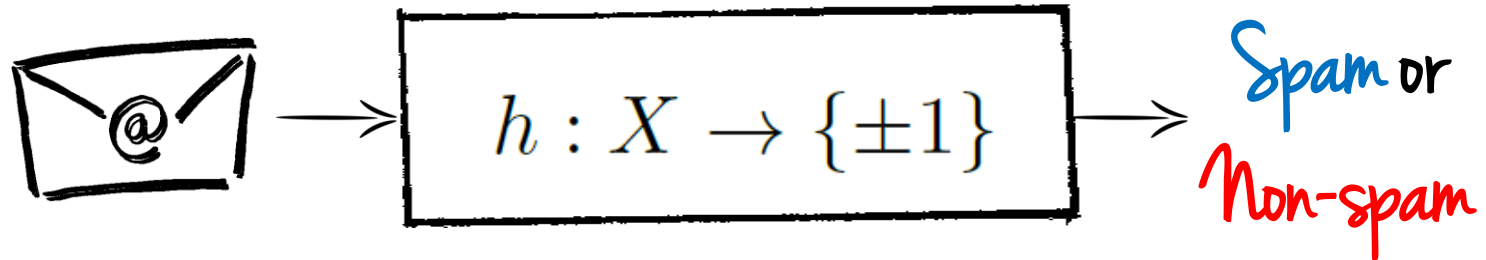


Binary Classification



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

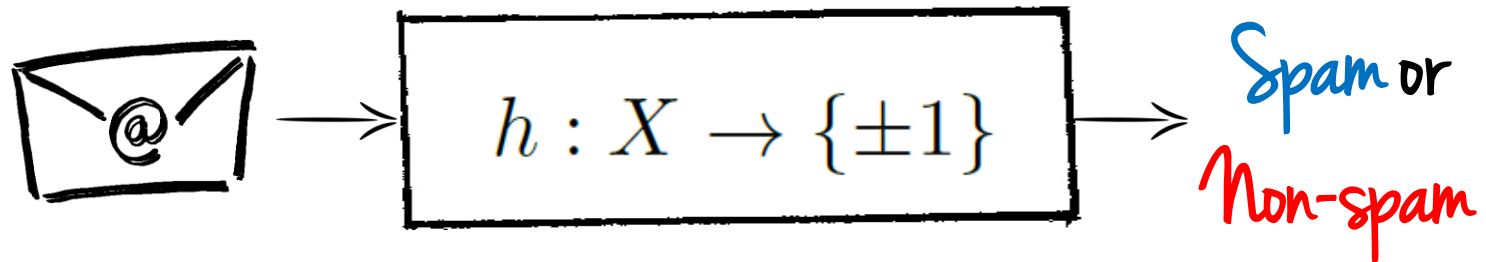
Binary Classification



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

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

Binary Classification

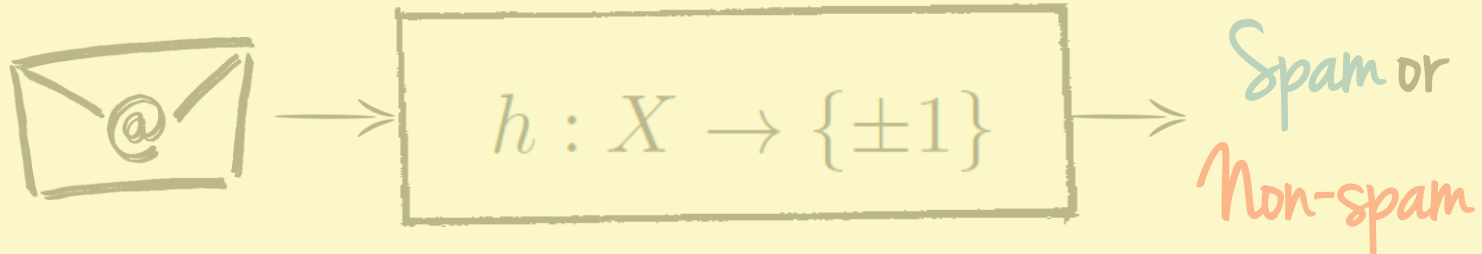


$$\text{er}_D^{0-1}[h] = \mathbf{E}_{(x,y) \sim D} [\mathbf{1}(h(x) \neq y)]$$

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

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

Binary Classification



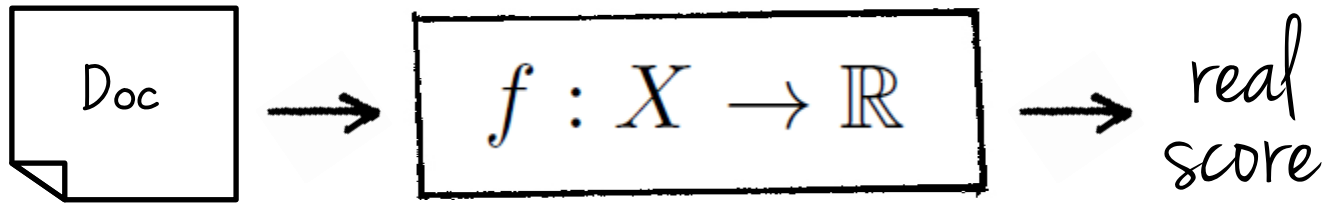
$S = ((x_1, y_1), \dots, (x_n, y_n))$ drawn iid from D

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

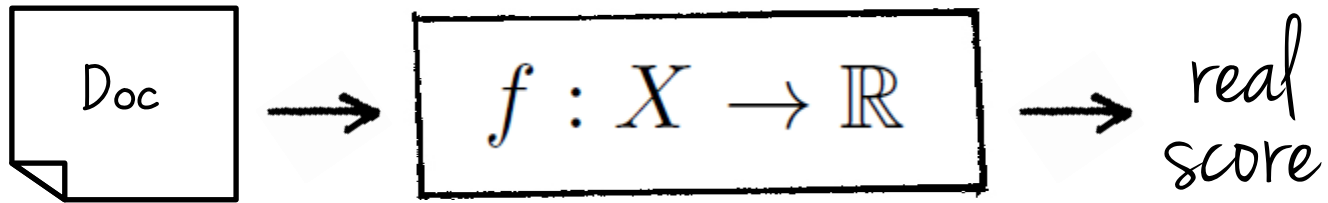
0-1 Consistency

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

Bipartite Ranking

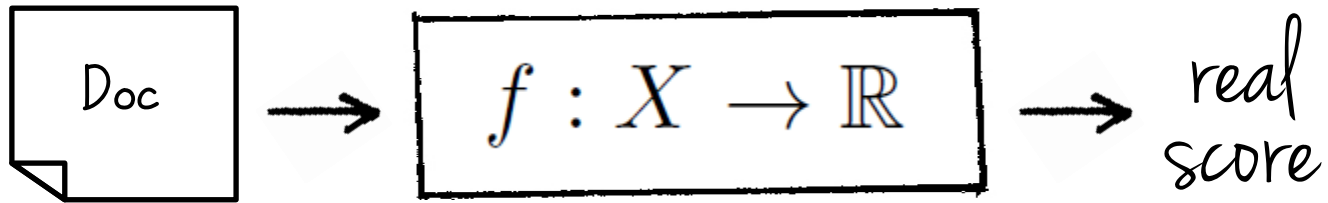


Bipartite Ranking



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

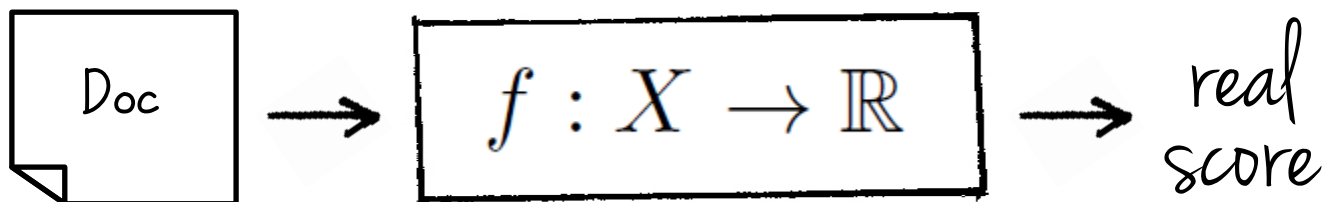
Bipartite Ranking



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

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

Bipartite Ranking

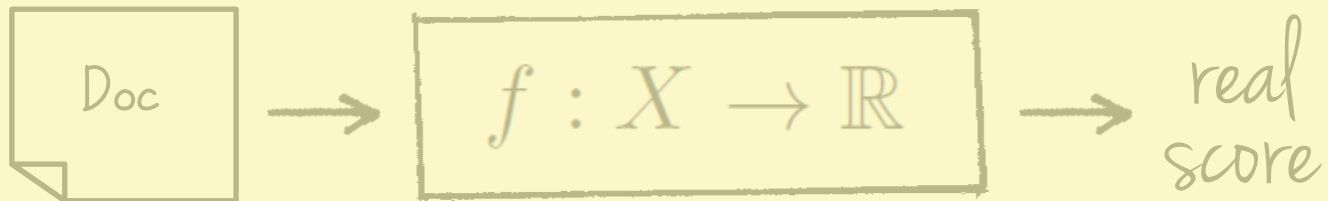


$$\text{er}_D^{\text{rank}}[f] = \mathbf{E} \left[\mathbf{1} \left((y - y')(f(x) - f(x')) < 0 \right) \mid y \neq y' \right]$$

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

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

Bipartite Ranking

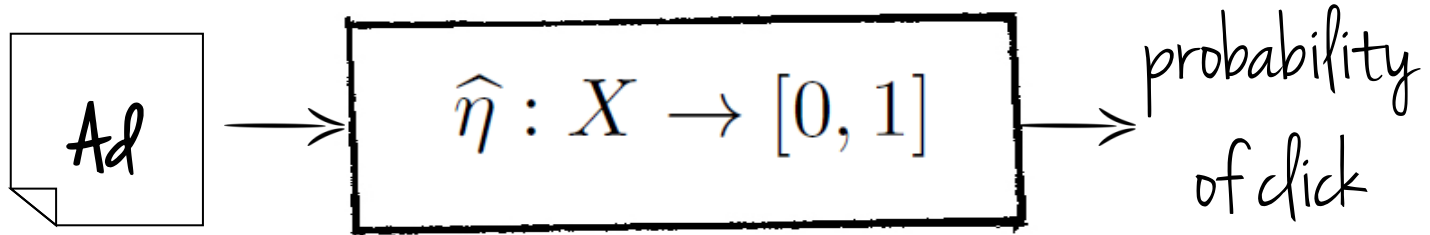


$S = ((x_1, y_1), \dots, (x_n, y_n))$ drawn iid from D

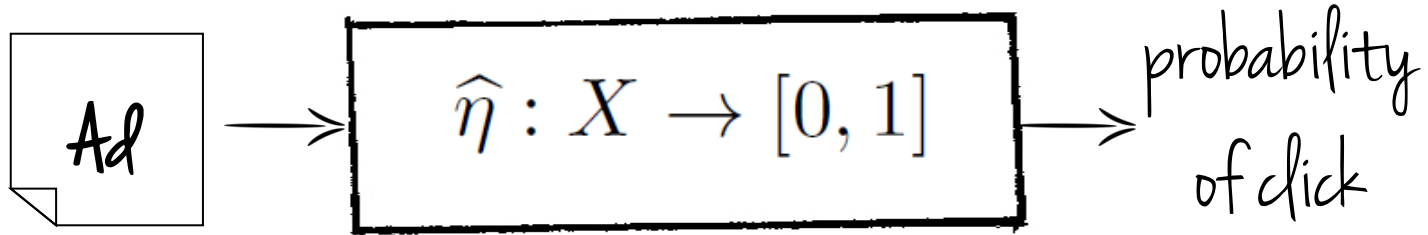
$$\boxed{\text{Ranking Consistency}} \\ \text{regret}_D^{\text{rank}}[f_S] \xrightarrow{P} 0$$

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

Binary Class Probability Estimation

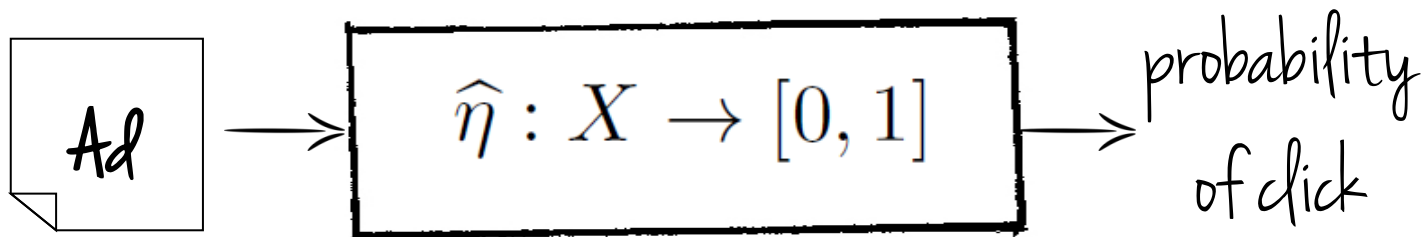


Binary Class Probability Estimation



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

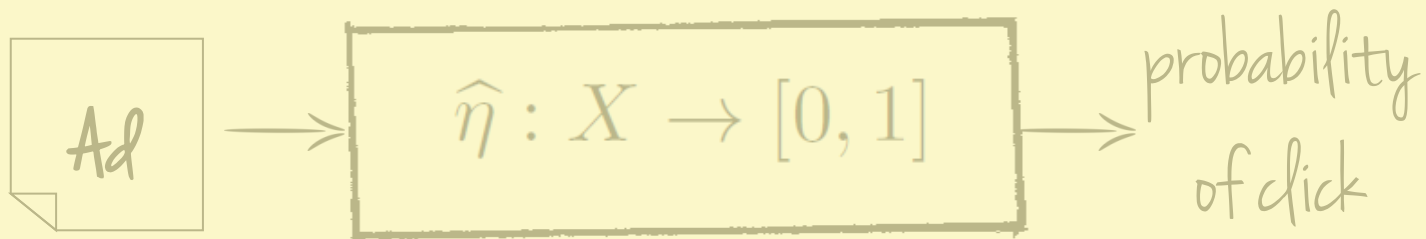
Binary Class Probability Estimation



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

$$\text{regret}_D^{\text{sq}}[\hat{\eta}] = \mathbf{E}_x [(\hat{\eta}(x) - \eta(x))^2]$$

Binary Class Probability Estimation



$S = ((x_1, y_1), \dots, (x_n, y_n))$ drawn iid from D

CPE Consistency

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

$$\text{regret}_D^{\text{sq}}[\hat{\eta}] = \mathbf{E}_x [(\hat{\eta}(x) - \eta(x))^2]$$

0-1 Consistency

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

Ranking Consistency

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

CPE Consistency

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

Statistical Consistency

(Surrogate Loss Minimization)

Statistical Consistency

(Surrogate Loss Minimization)

Loss	$\ell(y, \hat{y})$	Binary Classification	Bipartite Ranking*	Binary CPE
Logistic	$\ln(1 + e^{-y\hat{y}})$	✓	✓	✓
Exponential	$e^{-y\hat{y}}$	✓	✓	✓
Square	$(1 - y\hat{y})^2$	✓	✓	✓
Sq. Hinge	$((1 - y\hat{y})_+)^2$	✓	✓	✓
Hinge	$(1 - y\hat{y})_+$	✓	×	×

*pair-wise losses

Zhang (2004); Bartlett et al. (2006); Clemencon et al. (2008); Uematsu & Lee (2011); Agarwal (2012); Gao & Zhou (2013); Gao et al. (2013)

Ranking Consistency

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

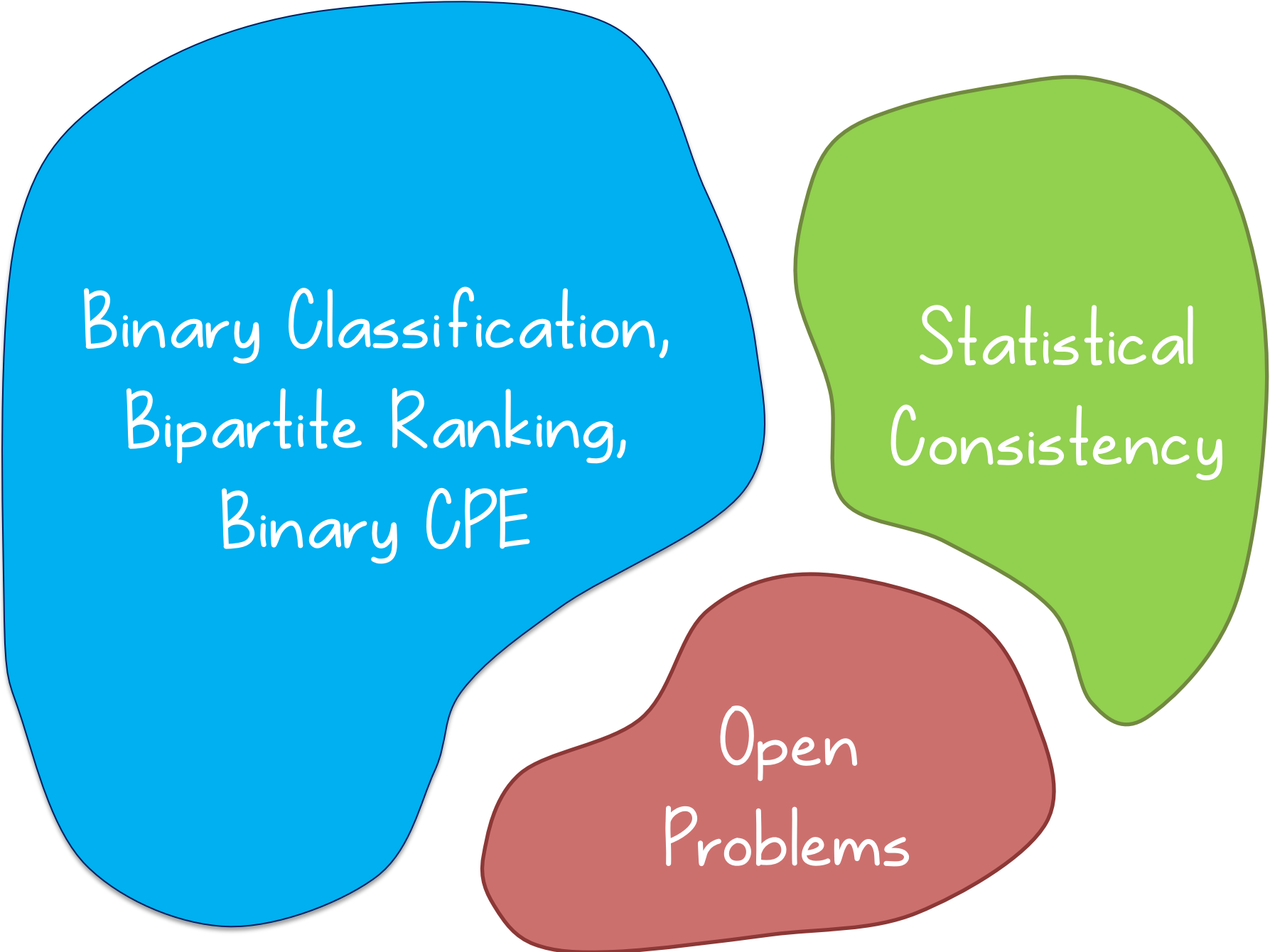
?

0-1 Consistency

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

CPE Consistency

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

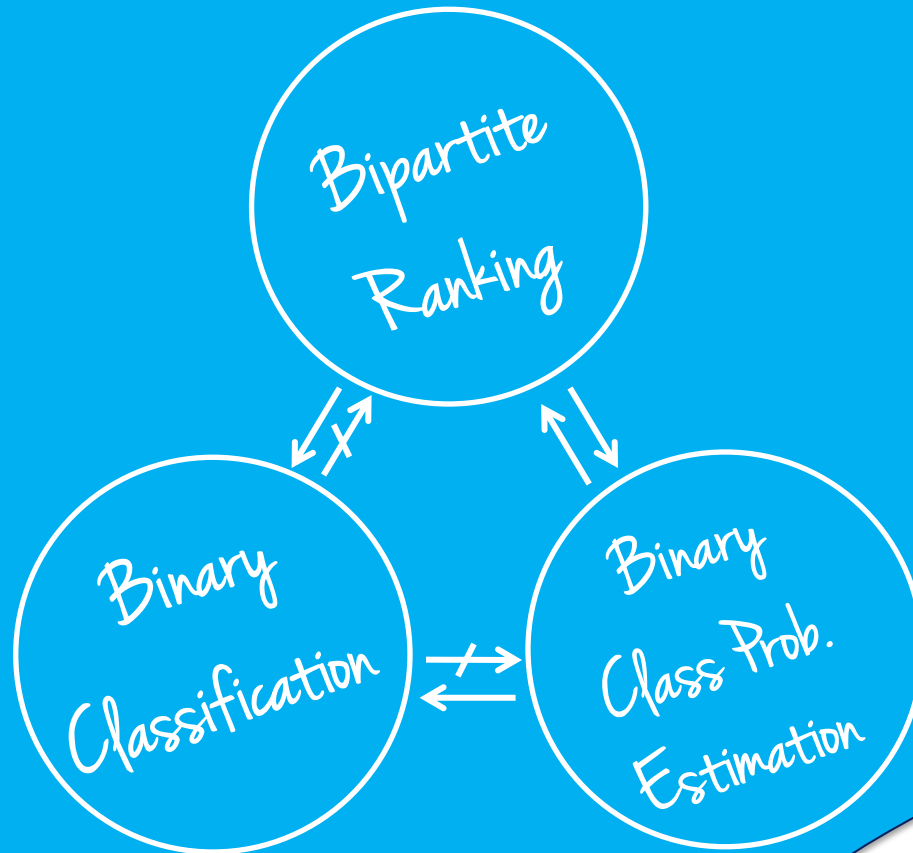


Binary Classification,
Bipartite Ranking,
Binary CPE

Statistical
Consistency

Open
Problems

Relationship Between BC, BR, BCPE

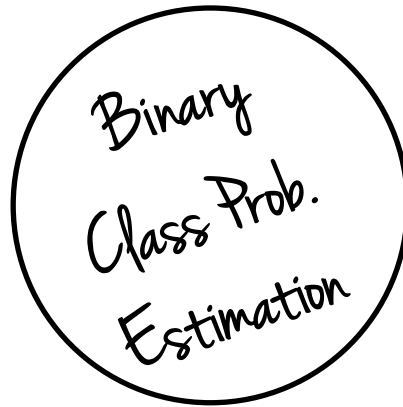


Bipartite
Ranking

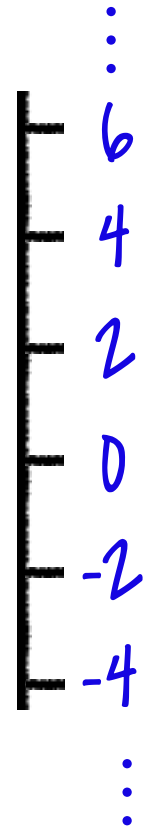
Binary
Classification

Binary
Class Prob.
Estimation





positive
or
negative



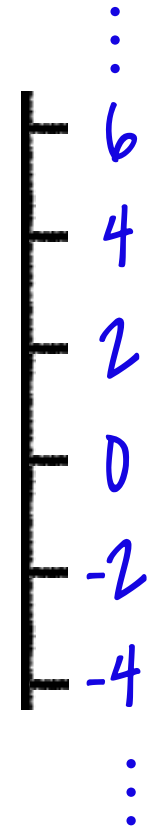
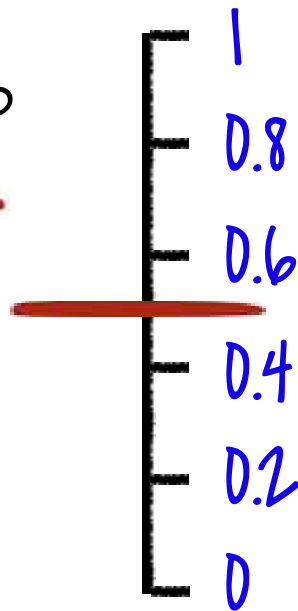
Binary
Classification

Binary
Class Prob.
Estimation

Bipartite
Ranking

positive
or
negative

threshold at
0.5



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

L. Devroye, L. Györfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition*. Springer, 1996.

C. Scott, “Calibrated asymmetric surrogate losses,” *Electronic Journal of Statistics*, vol. 6, pp. 958–992, 2012.

$$\text{regret}_D^{0-1}[\text{sign} \circ (\hat{\eta} - \tfrac{1}{2})] \leq \sqrt{\text{regret}_D^{\text{sq}}[\hat{\eta}]} \xrightarrow{P} 0$$

L. Devroye, L. Györfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition*. Springer, 1996.

C. Scott, “Calibrated asymmetric surrogate losses,” *Electronic Journal of Statistics*, vol. 6, pp. 958–992, 2012.

$$\boxed{\text{regret}_D^{0-1} [\text{sign} \circ (\hat{\eta} - \tfrac{1}{2})]} \leq \boxed{\sqrt{\text{regret}_D^{\text{sq}} [\hat{\eta}]}}$$

$\xrightarrow{P} 0$

$\xrightarrow{P} 0$

L. Devroye, L. Györfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition*. Springer, 1996.

C. Scott, “Calibrated asymmetric surrogate losses,” *Electronic Journal of Statistics*, vol. 6, pp. 958–992, 2012.

Regret Transfer Bound

$$\boxed{\text{regret}_D^{0-1}[\text{sign} \circ (\hat{\eta} - \tfrac{1}{2})]} \leq \boxed{\sqrt{\text{regret}_D^{\text{sq}}[\hat{\eta}]}}$$

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

L. Devroye, L. Györfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition*. Springer, 1996.

C. Scott, “Calibrated asymmetric surrogate losses,” *Electronic Journal of Statistics*, vol. 6, pp. 958–992, 2012.

Binary
Classification

Binary
Class Prob.
Estimation

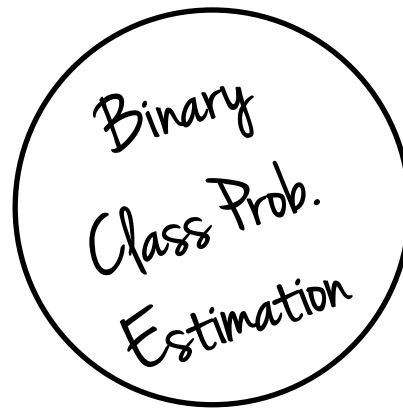
Bipartite
Ranking

positive
or
negative

threshold at
0.5
←

1
0.8
0.6
0.4
0.2
0

⋮
6
4
2
0
-2
-4
⋮

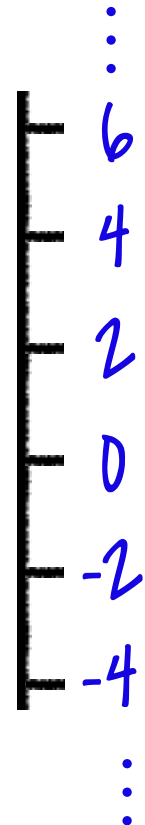


positive
or
negative

threshold at
0.5
←



rank by
probabilities
→



Regret Transfer Bound

$$\text{regret}_D^{\text{rank}}[\hat{\eta}] \leq C \sqrt{\text{regret}_D^{\text{sq}}[\hat{\eta}]}$$

Regret Transfer Bound

$$\text{regret}_D^{\text{rank}}[\hat{\eta}] \leq C \sqrt{\text{regret}_D^{\text{sq}}[\hat{\eta}]}$$

$\xrightarrow{P} 0$

Regret Transfer Bound

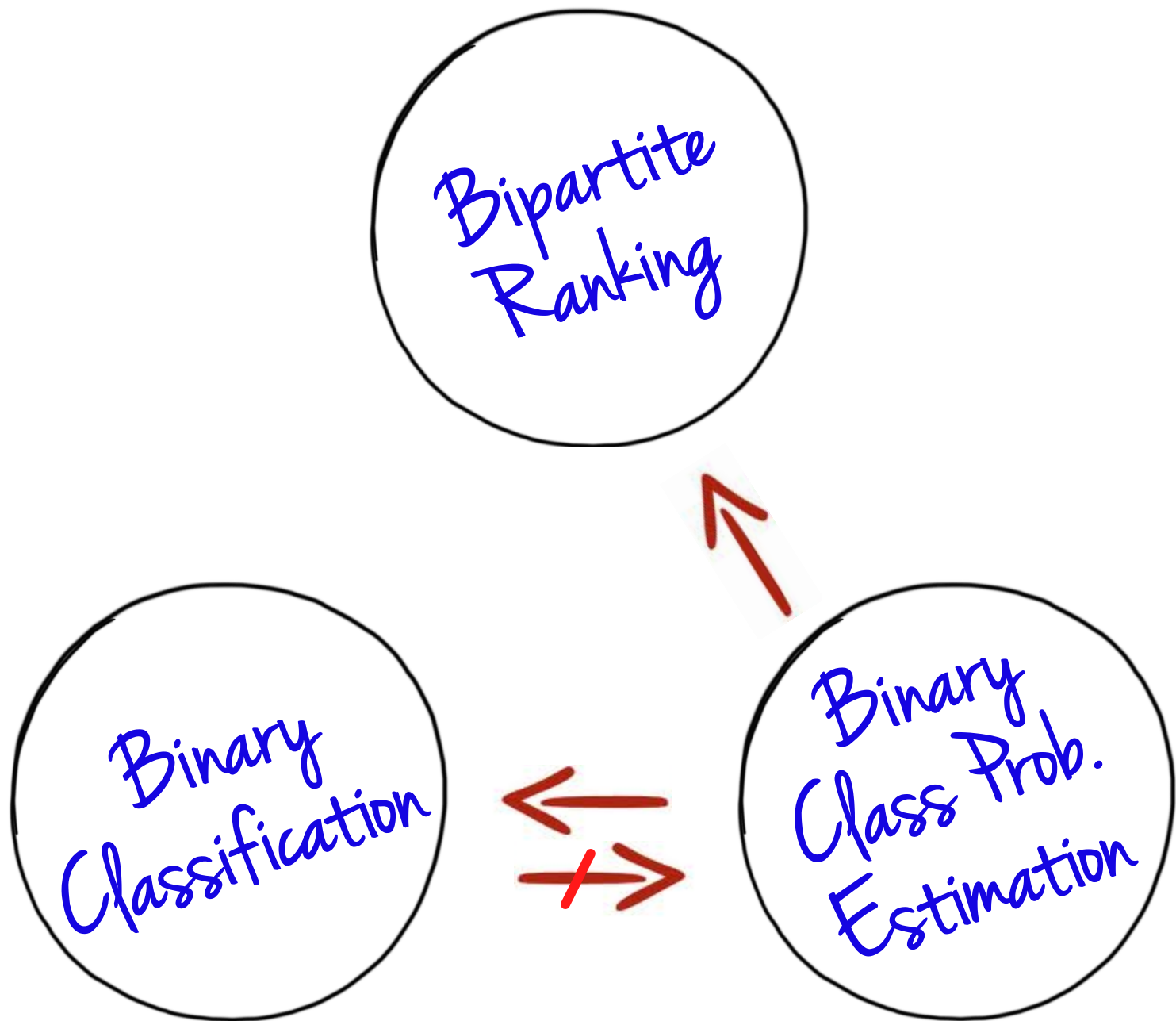
$$\boxed{\text{regret}_D^{\text{rank}}[\hat{\eta}]}_{\xrightarrow{P} 0} \leq C \sqrt{\boxed{\text{regret}_D^{\text{sq}}[\hat{\eta}]}_{\xrightarrow{P} 0}}$$

Bipartite
Ranking

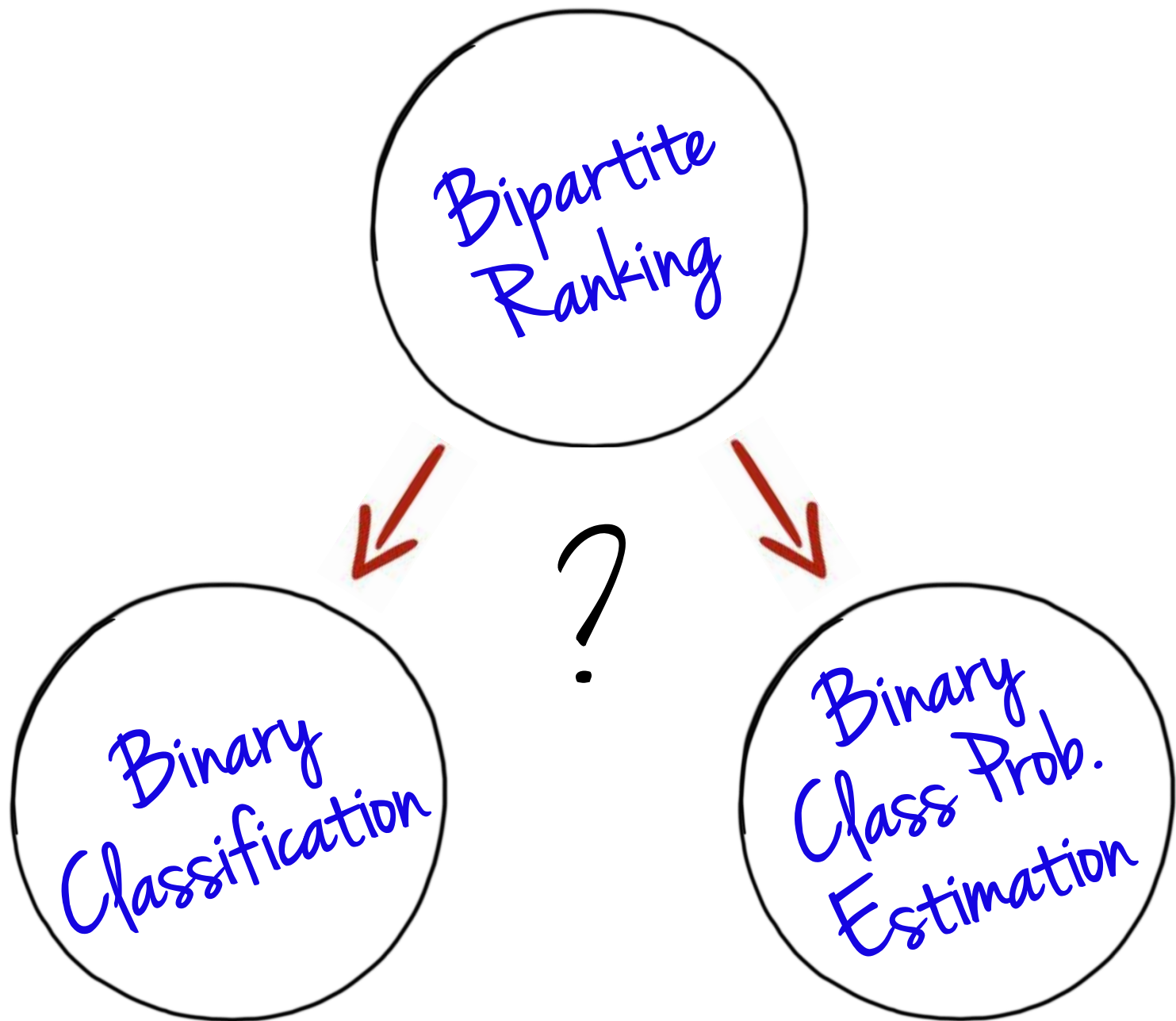
Binary
Classification

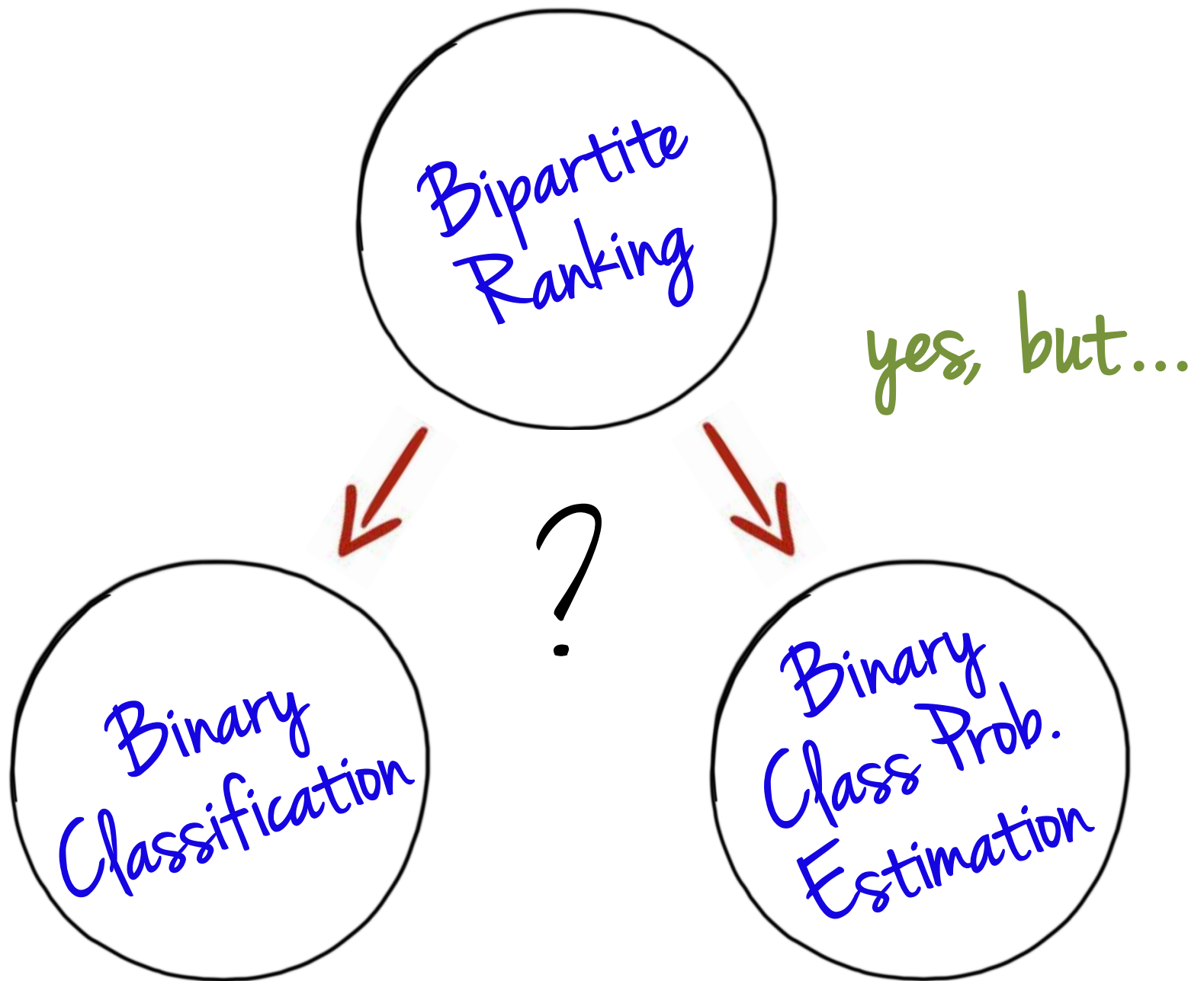
Binary
Class Prob.
Estimation

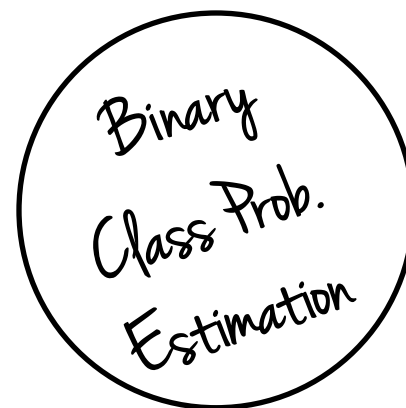




T. Zhang, "Statistical behavior and consistency of classification methods based on convex risk minimization," *Annals of Statistics*, vol. 32, no. 1, pp. 56–85, 2004.







positive
or
negative



Bipartite Ranking $\rightarrow$ Binary Classification

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

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

Bipartite Ranking $\rightarrow$ Binary Classification

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

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

Regret Transfer Bound

$$\operatorname{regret}_D^{0-1} [\operatorname{sign} \circ (f - t^*)] \leq C' \sqrt{\operatorname{regret}_D^{\operatorname{rank}}[f]}$$

Bipartite Ranking $\rightarrow$ Binary Classification

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

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

Regret Transfer Bound

$$\operatorname{regret}_D^{0-1} [\operatorname{sign} \circ (f - t^*)] \leq C' \sqrt{\operatorname{regret}_D^{\operatorname{rank}}[f]} \rightarrow 0$$

Bipartite Ranking $\rightarrow$ Binary Classification

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

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

Regret Transfer Bound

$$\boxed{\operatorname{regret}_D^{0-1} [\operatorname{sign} \circ (f - t^*)]} \leq C' \sqrt{\boxed{\operatorname{regret}_D^{\operatorname{rank}} [f]}}$$

$\rightarrow 0 \qquad \qquad \qquad \rightarrow 0$

Bipartite Ranking $\rightarrow$ Binary Classification

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

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

(Weak) Regret Transfer Bound

$$\boxed{\operatorname{regret}_D^{0-1} [\operatorname{sign} \circ (f - t^*)]} \leq C' \sqrt{\boxed{\operatorname{regret}_D^{\operatorname{rank}} [f]}}$$

dependence on distribution $\rightarrow 0$ $\rightarrow 0$

Bipartite Ranking $\rightarrow$ Binary Classification

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

$S = ((x_1, y_1), \dots, (x_n, y_n))$ drawn iid from D

$$\hat{t} \in \operatorname{argmax}_{t \in \mathbb{R}} \left\{ \operatorname{er}_S^{0-1} [\operatorname{sign} \circ (f - t)] \right\}$$

Sample-Based (Weak) Regret Transfer Bound

$$\operatorname{regret}_D^{0-1} [\operatorname{sign} \circ (f - \hat{t})] \leq C' \sqrt{\operatorname{regret}_D^{\operatorname{rank}}[f]} + O\left(\sqrt{\frac{\ln(n)}{n}}\right)$$

Bipartite Ranking $\rightarrow$ Binary Classification

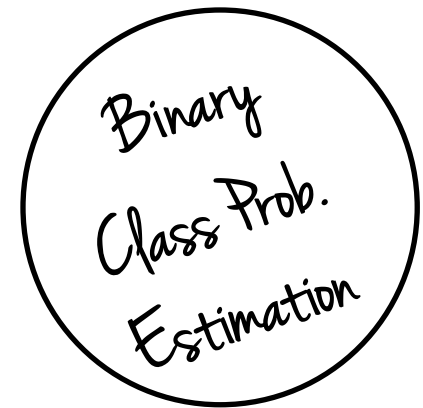
$$f : X \rightarrow \mathbb{R}$$

$S = ((x_1, y_1), \dots, (x_n, y_n))$ drawn iid from D

$$\hat{t} \in \operatorname{argmax}_{t \in \mathbb{R}} \left\{ \operatorname{er}_S^{0-1} [\operatorname{sign} \circ (f - t)] \right\}$$

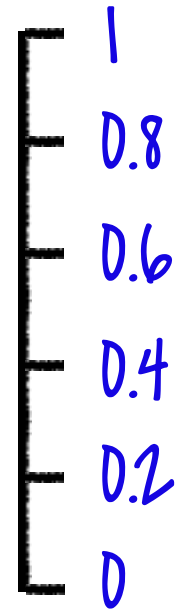
Sample-Based (Weak) Regret Transfer Bound

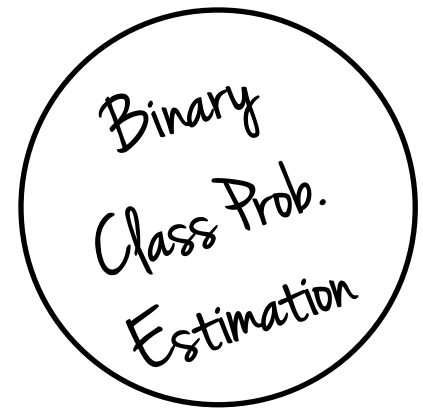
$$\operatorname{regret}_D^{0-1} [\operatorname{sign} \circ (f - \hat{t})] \leq C' \sqrt{\operatorname{regret}_D^{\operatorname{rank}}[f]} + \boxed{O\left(\sqrt{\frac{\ln(n)}{n}}\right)} \rightarrow 0$$



positive
or
negative

additional
training data
←
Optimal
threshold





positive
or
negative

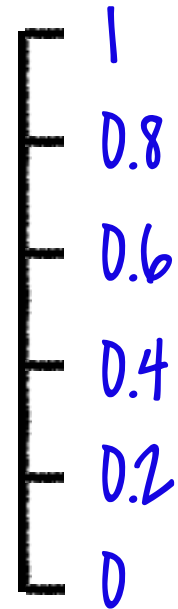
additional
training data

←
Optimal
threshold



additional
training data

→
Calibration



Bipartite Ranking $\rightarrow$ Binary CPE

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

$$\text{Cal} \in \operatorname{argmax}_{g \in \mathcal{G}_{\text{inc}}} \left\{ \operatorname{er}_D^{\text{sq}} [g \circ f] \right\}$$

 **monotonically increasing functions**

Bipartite Ranking $\rightarrow$ Binary CPE

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

$$\text{Cal} \in \operatorname{argmax}_{g \in \mathcal{G}_{\text{inc}}} \left\{ \operatorname{er}_D^{\text{sq}}[g \circ f] \right\}$$

$\swarrow$ monotonically increasing functions

(Weak) Regret Transfer Bound

$$\operatorname{regret}_D^{\text{sq}}[\text{Cal} \circ f] \leq C'' \sqrt{\operatorname{regret}_D^{\text{rank}}[f]}$$

Bipartite Ranking $\rightarrow$ Binary CPE

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

$$\text{Cal} \in \operatorname{argmax}_{g \in \mathcal{G}_{\text{inc}}} \left\{ \operatorname{er}_D^{\text{sq}}[g \circ f] \right\}$$

$\swarrow$
monotonically increasing functions

(Weak) Regret Transfer Bound

$$\boxed{\operatorname{regret}_D^{\text{sq}}[\text{Cal} \circ f]} \leq C'' \sqrt{\boxed{\operatorname{regret}_D^{\text{rank}}[f]}}$$

$\rightarrow 0 \qquad \qquad \qquad \rightarrow 0$

Bipartite Ranking $\rightarrow$ Binary CPE

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

$$\text{Cal} \in \operatorname{argmax}_{g \in \mathcal{G}_{\text{inc}}} \left\{ \operatorname{er}_D^{\text{sq}}[g \circ f] \right\}$$

$\swarrow$ monotonically increasing functions

(Weak) Regret Transfer Bound

$$\boxed{\operatorname{regret}_D^{\text{sq}}[\text{Cal} \circ f]} \leq C'' \sqrt{\boxed{\operatorname{regret}_D^{\text{rank}}[f]}}$$

$\rightarrow 0 \qquad \qquad \qquad \rightarrow 0$

dependence on distribution $\swarrow$

Bipartite Ranking $\rightarrow$ Binary CPE

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

$S = ((x_1, y_1), \dots, (x_n, y_n))$ drawn iid from D

$$\widehat{\text{Cal}} \in \operatorname{argmax}_{g \in \mathcal{G}_{\text{inc}}} \left\{ \operatorname{er}_S^{\text{sq}} [g \circ f] \right\}$$

Sample-Based (Weak) Regret Transfer Bound

$$\operatorname{regret}_D^{\text{sq}} [\widehat{\text{Cal}} \circ f] \leq C'' \sqrt{\operatorname{regret}_D^{\text{rank}} [f]} + O\left(\left(\frac{\ln(n)}{n}\right)^{1/3}\right)$$

Bipartite Ranking $\rightarrow$ Binary CPE

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

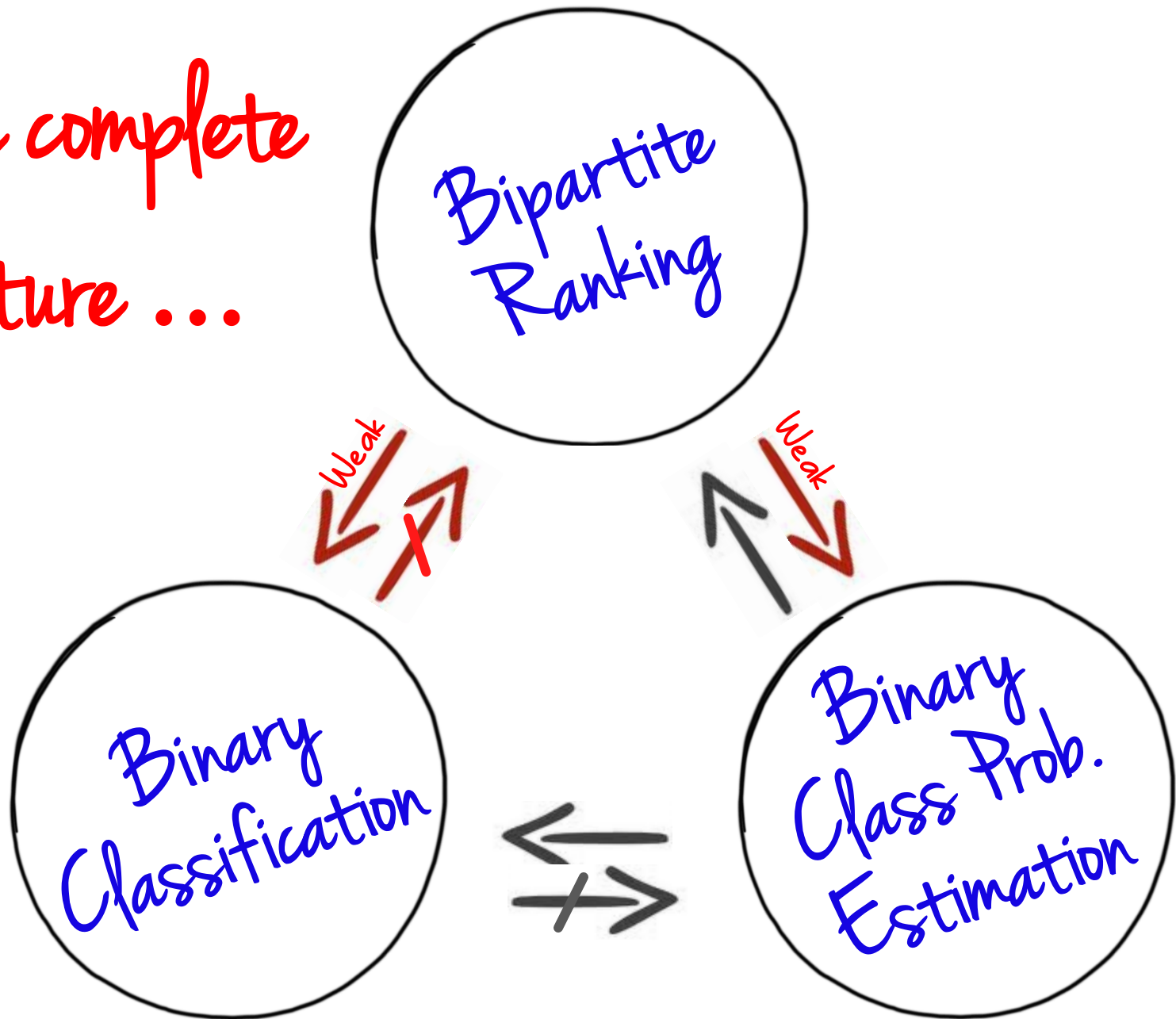
$S = ((x_1, y_1), \dots, (x_n, y_n))$ drawn iid from D

$$\widehat{\text{Cal}} \in \operatorname{argmax}_{g \in \mathcal{G}_{\text{inc}}} \left\{ \operatorname{er}_S^{\text{sq}} [g \circ f] \right\}$$

Sample-Based (Weak) Regret Transfer Bound

$$\operatorname{regret}_D^{\text{sq}} [\widehat{\text{Cal}} \circ f] \leq C'' \sqrt{\operatorname{regret}_D^{\text{rank}} [f]} + \boxed{O\left(\left(\frac{\ln(n)}{n}\right)^{1/3}\right)} \rightarrow 0$$

the complete
picture ...

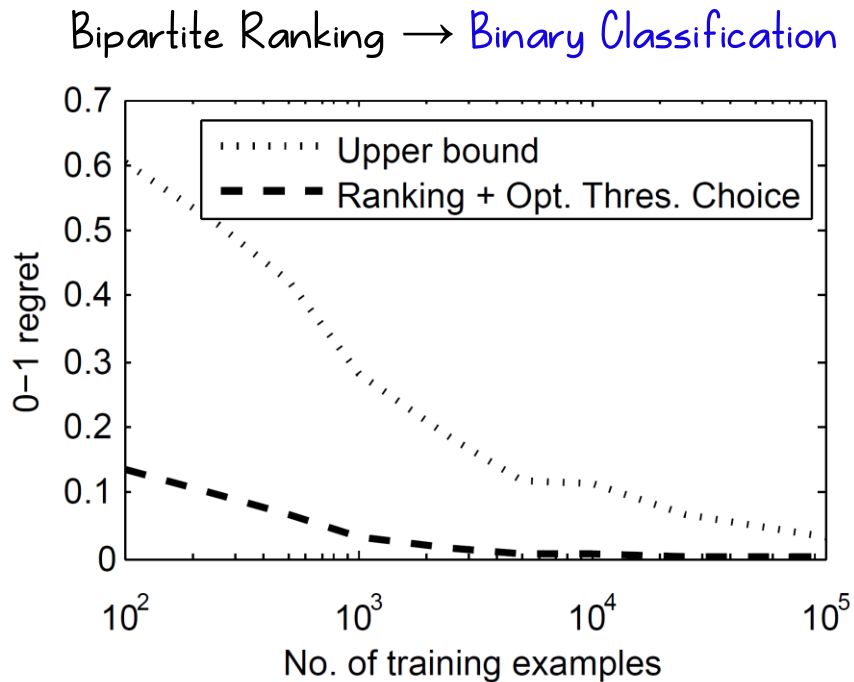


Experiments: Synthetic Data

Validating Regret Transfer Bounds

Experiments: Synthetic Data

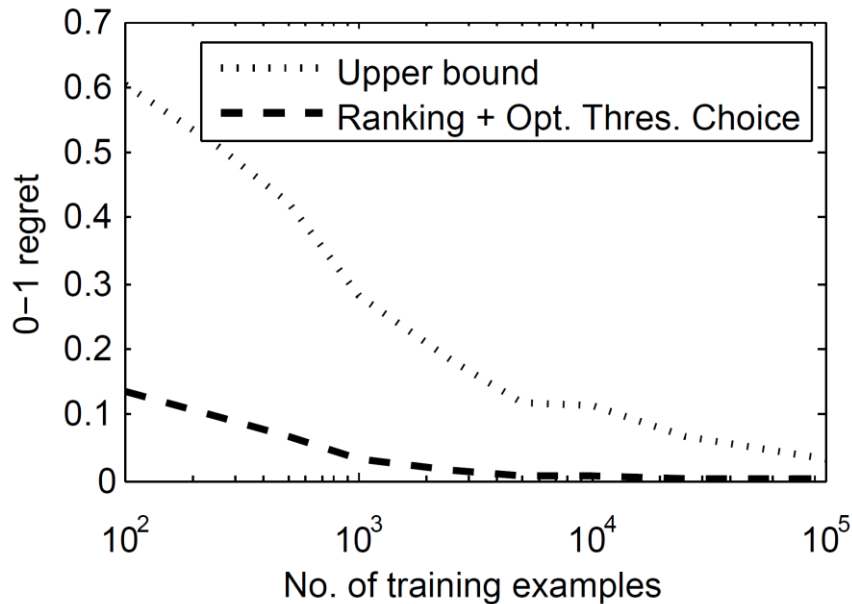
Validating Regret Transfer Bounds



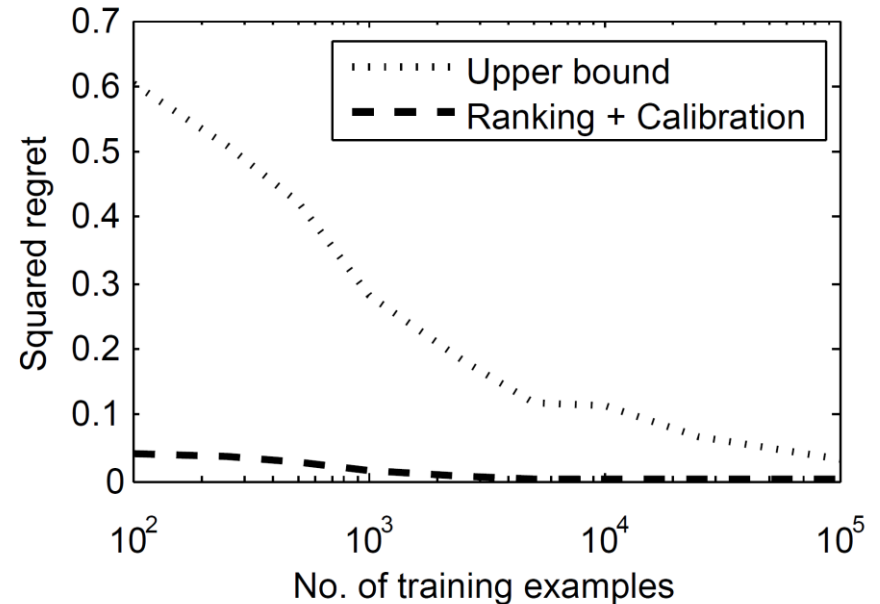
Experiments: Synthetic Data

Validating Regret Transfer Bounds

Bipartite Ranking $\rightarrow$ Binary Classification



Bipartite Ranking $\rightarrow$ Binary CPE



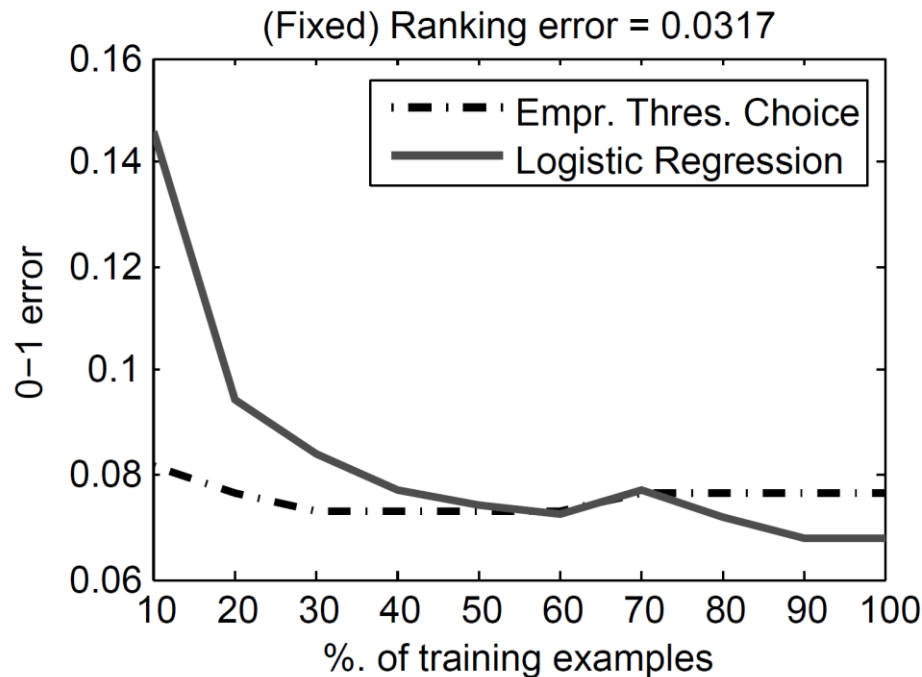
Experiments: Real Data

Comparison with Learning from Scratch

Experiments: Real Data

Comparison with Learning from Scratch

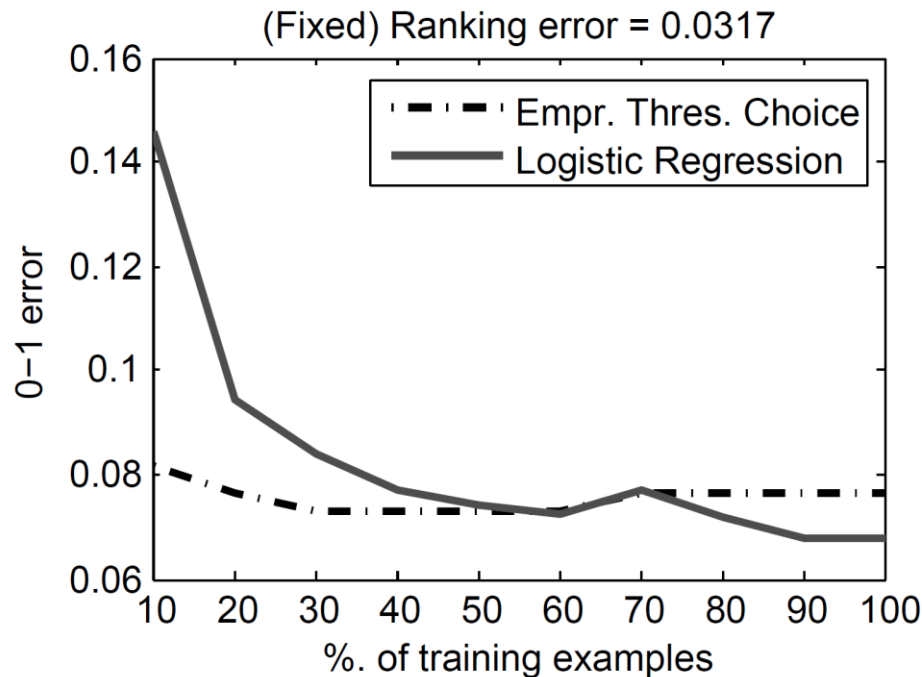
Bipartite Ranking $\rightarrow$ Binary Classification



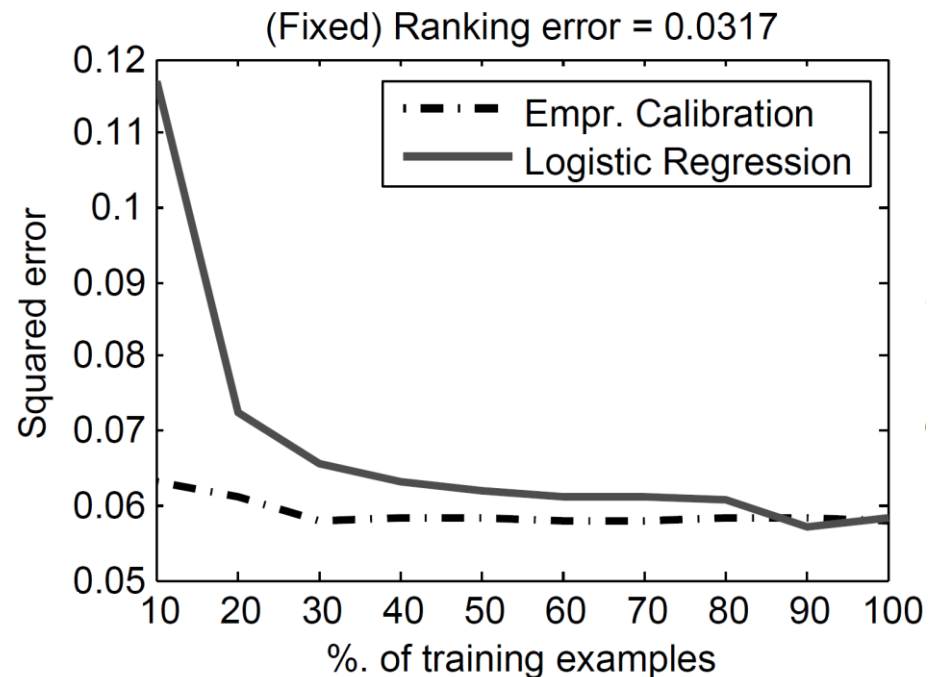
Experiments: Real Data

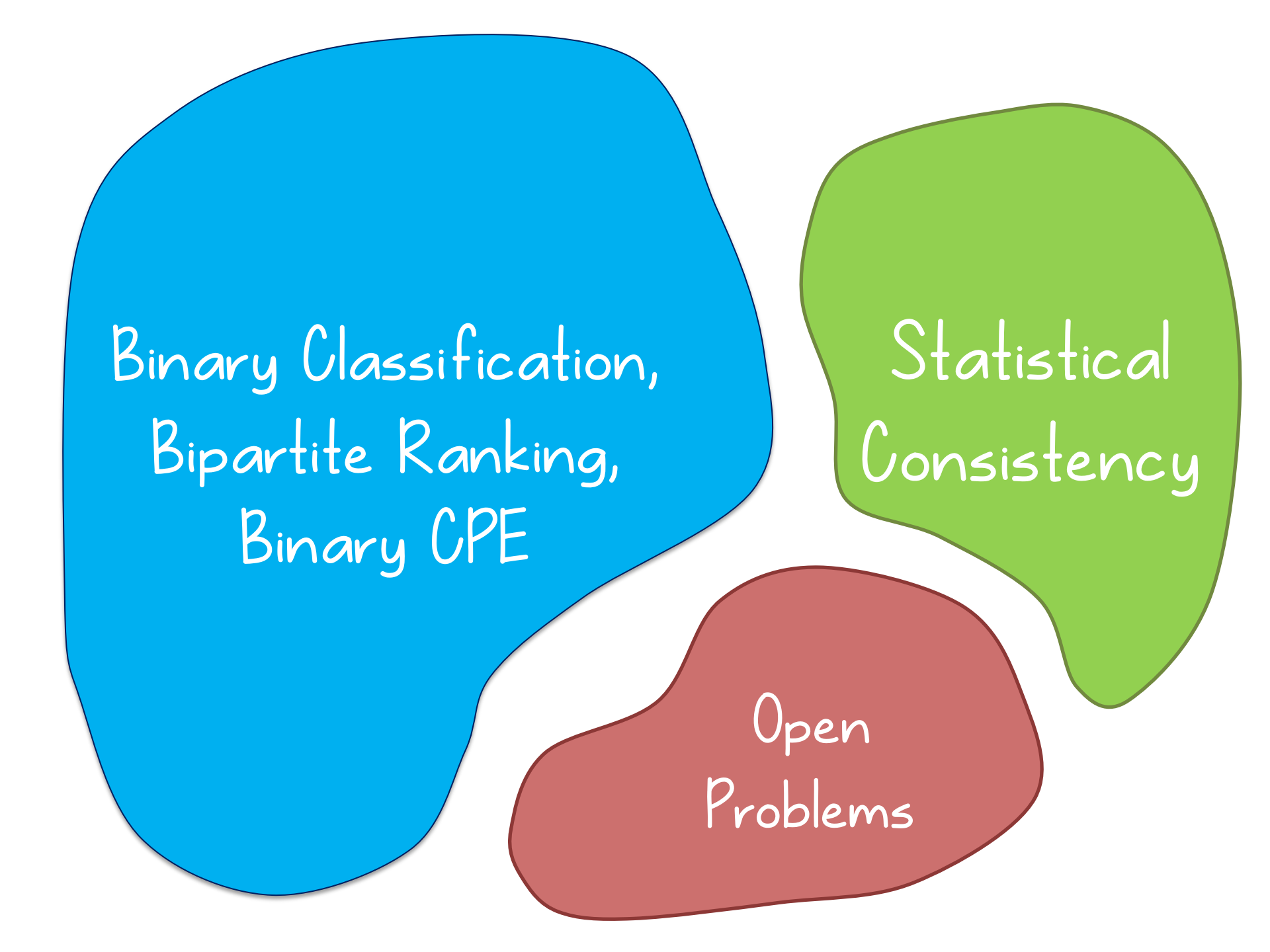
Comparison with Learning from Scratch

Bipartite Ranking $\rightarrow$ Binary Classification



Bipartite Ranking $\rightarrow$ Binary CPE



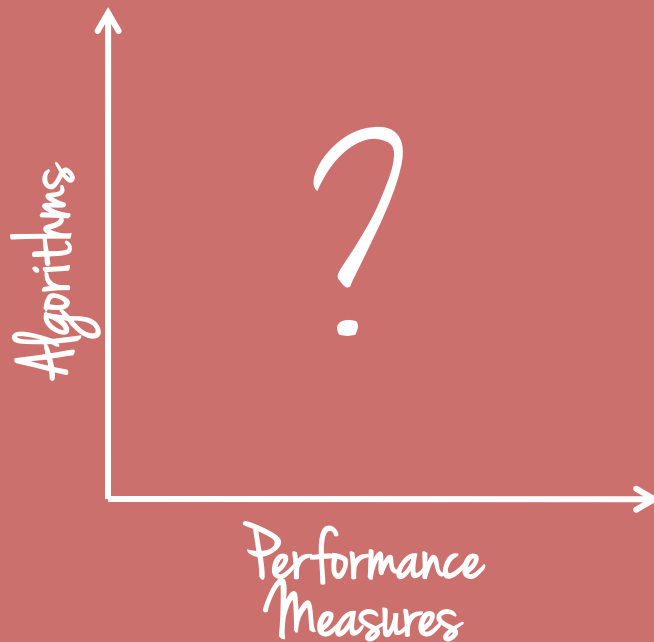


Binary Classification,
Bipartite Ranking,
Binary CPE

Statistical
Consistency

Open
Problems

Open Problems



performance measure?

Binary
Classification

0-1

Classification
Error

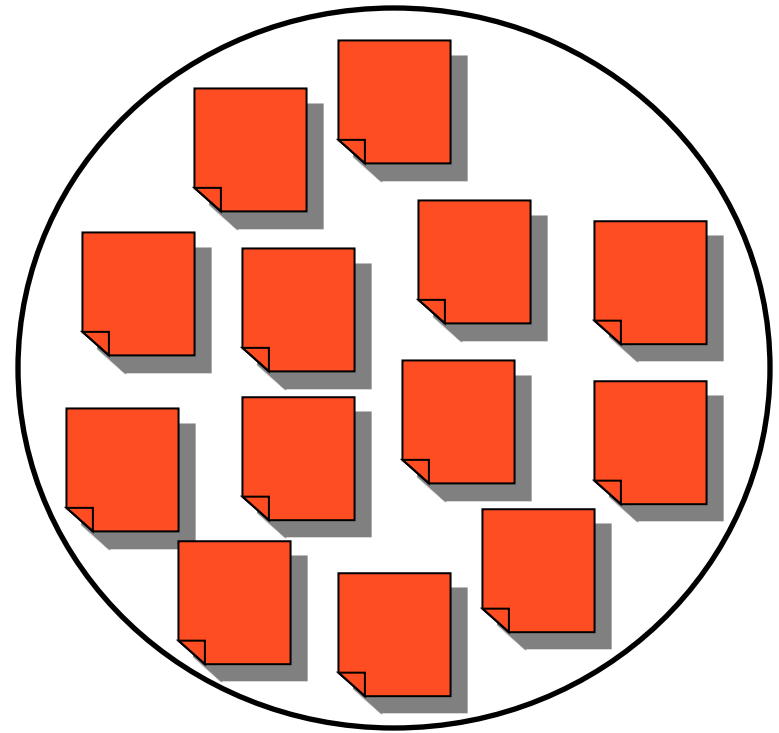
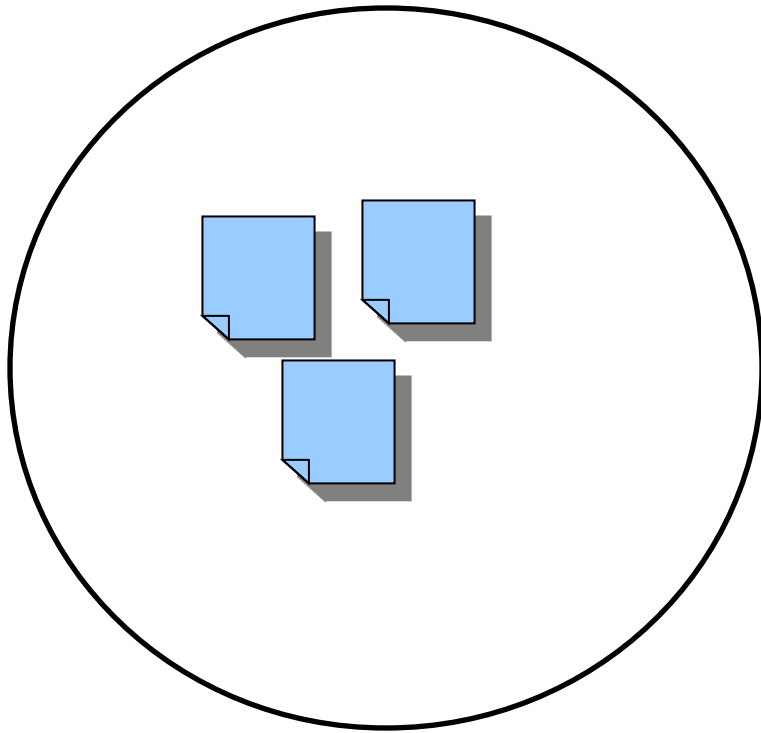
Bipartite
Ranking

Pair-wise
Accuracy

Binary
Class Prob.
Estimation

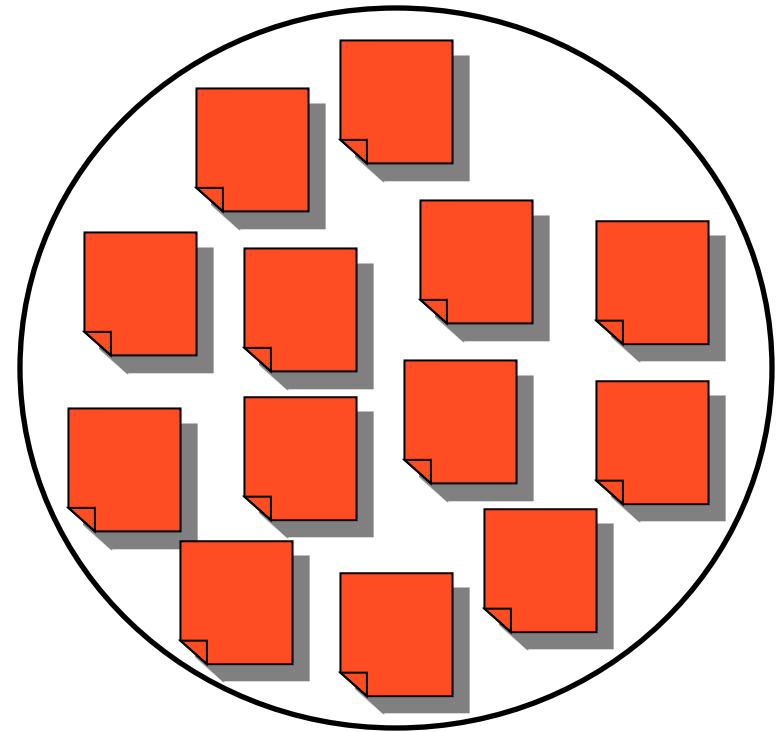
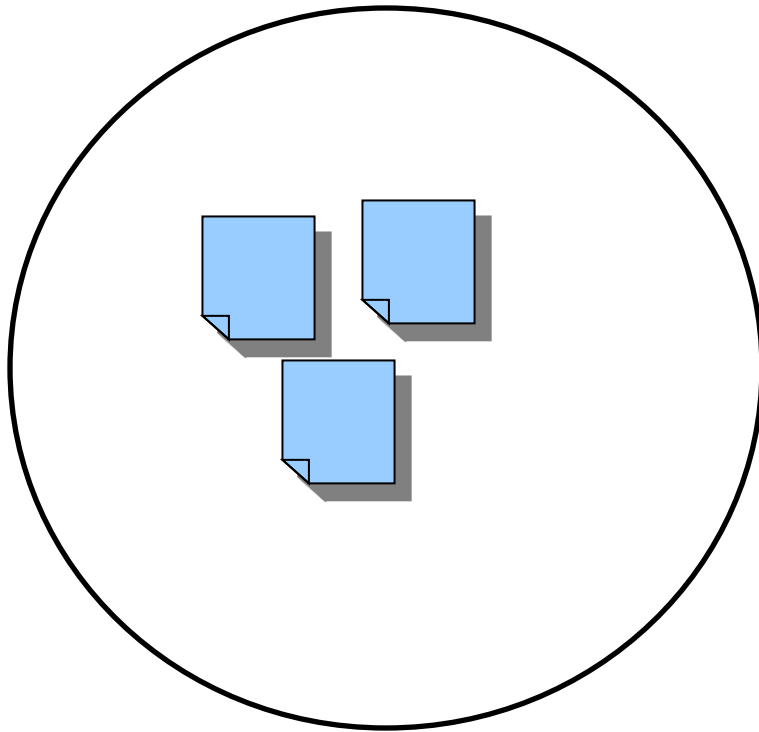
Squared
Error

Text Retrieval



Text Retrieval

F-measure



Information Retrieval



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](#)

Information Retrieval



learning to rank



Search

About 216,000,000 results (0.23 seconds)

Average Precision

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](#)

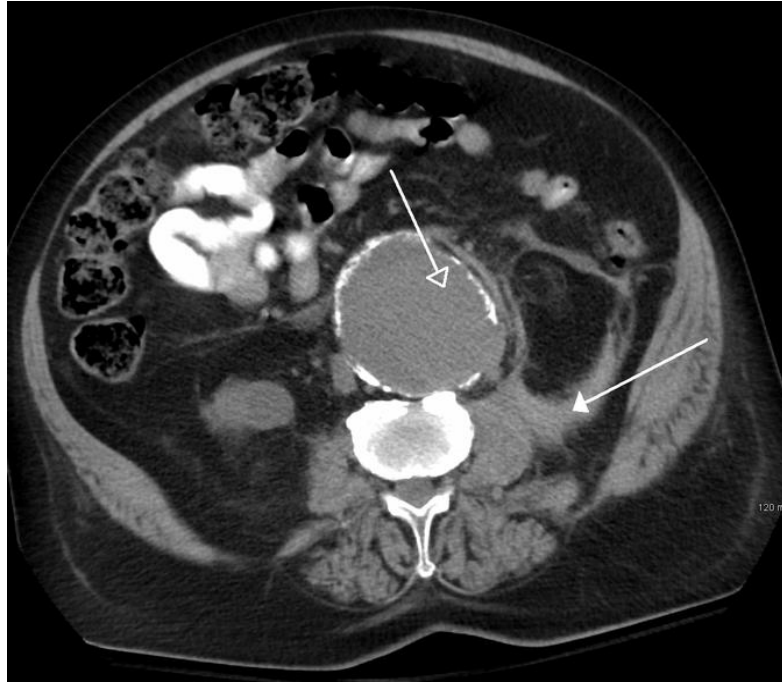
Learning to rank for information retrieval has gained a lot of interest in the ... field in which machine learning algorithms are used to learn this ranking function.

[PDF] [Future directions in learning to rank](#)

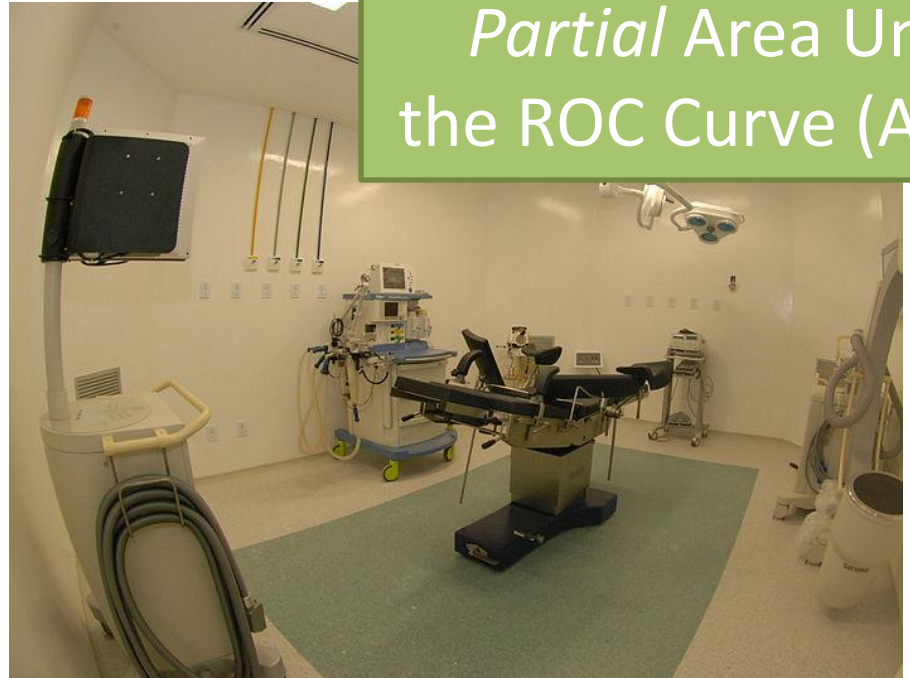
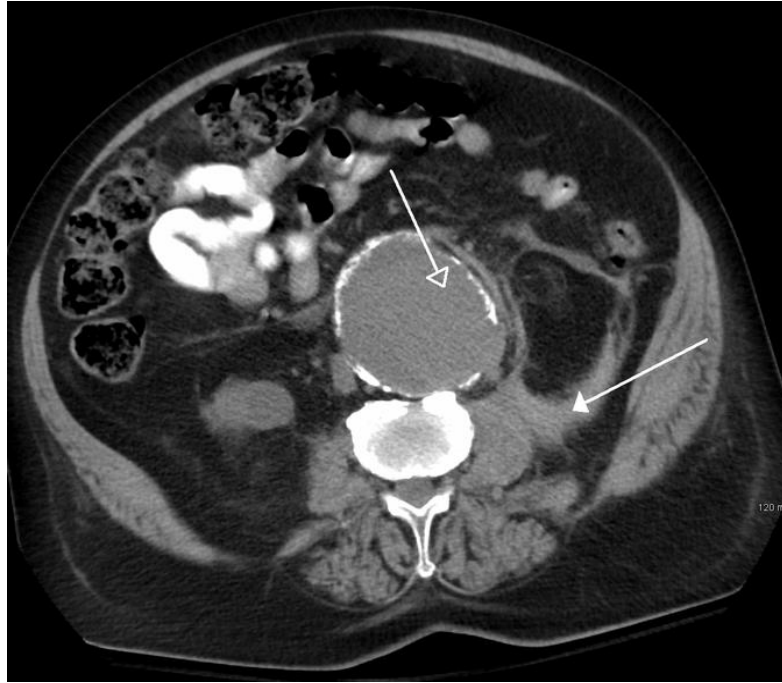
jmlr.csail.mit.edu/proceedings/papers/v14/.../chapelle11b.pdf

File Format: PDF/Adobe Acrobat - [Quick View](#)

Medical Diagnosis



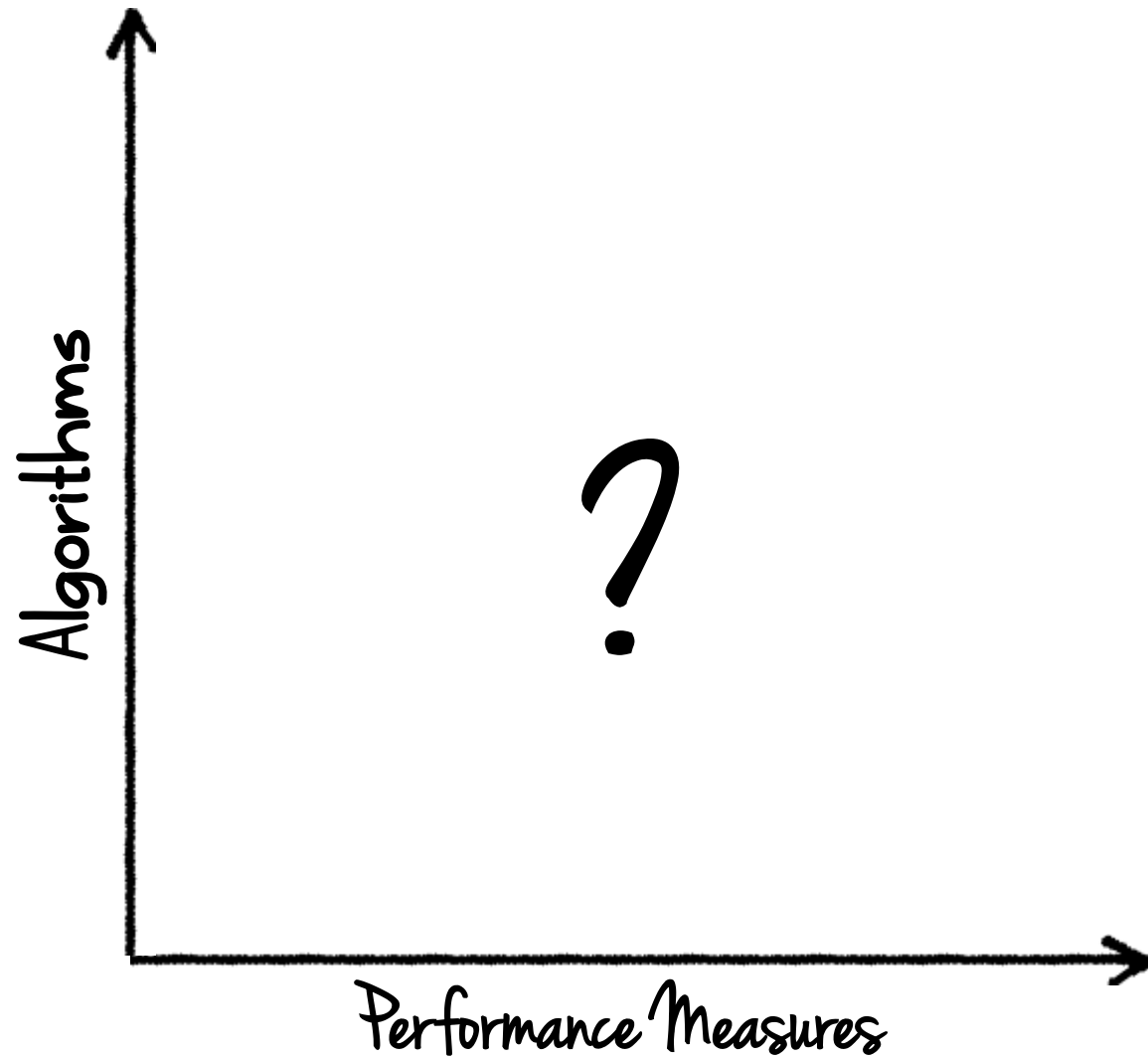
Medical Diagnosis



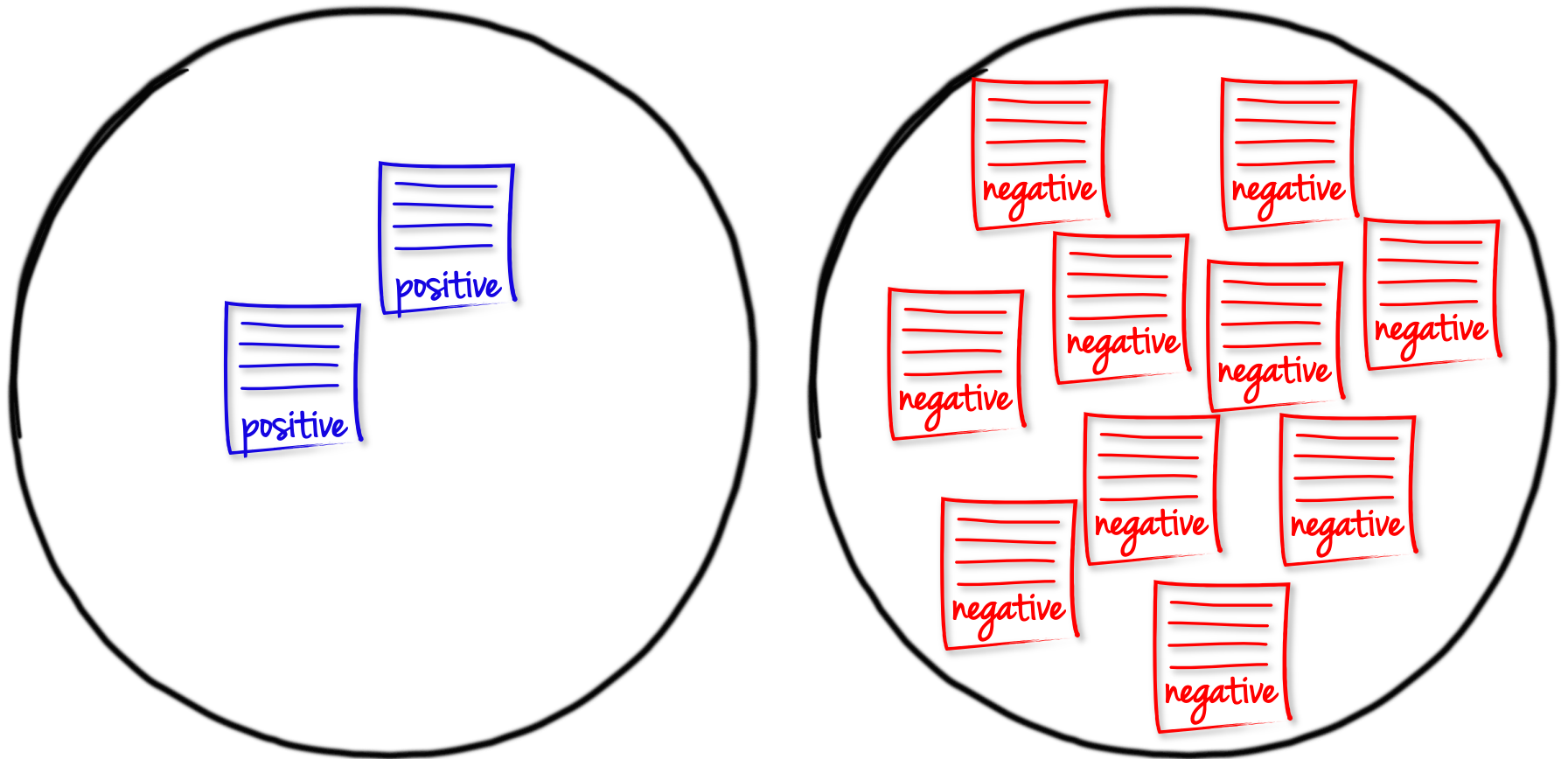
Partial Area Under the ROC Curve (AUC)



WTAF-measureAUCCalibrationLossLogLoss
ParialAUCAccuracy@TopPRBEP MRR
AM-measureg-mean
Accuracy@Top MRR StratifiedBS BalancedMPR
PRBEP
MAE MPR
StratifiedBS
Recall MPR
BalancedMPR
AUC
AM-measure
TPR
LogLoss
BrierScore
LocalAUC
FPR
RefinementLoss
F-measure
WeightedAccuracy
TPR
Positives@Top
BrierScore
Precision
g-mean
LocalAUC
MAE
Recall
AveragePrecision
Positives@Top
Accuracy
FPR

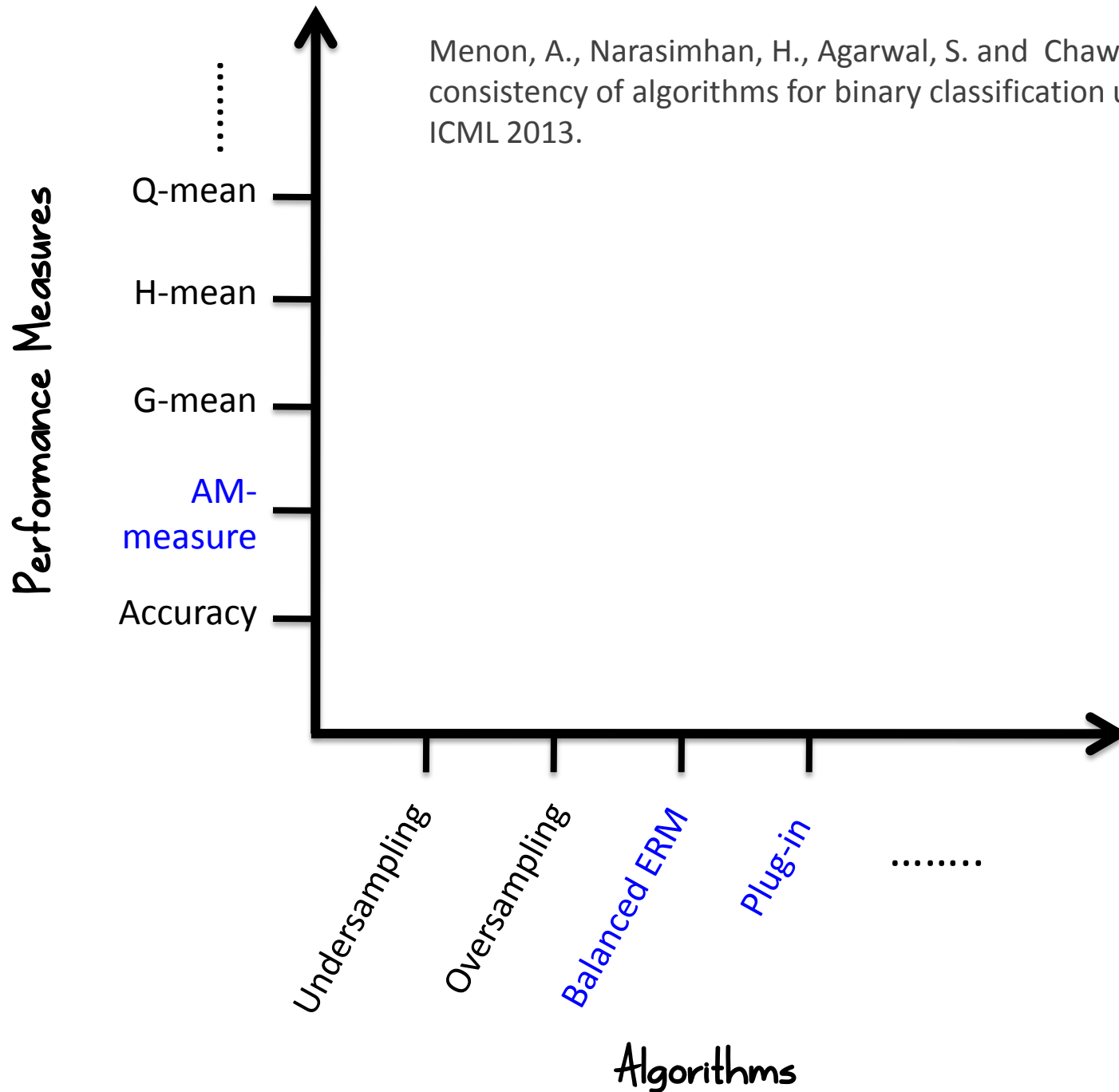


Example: Class Imbalanced

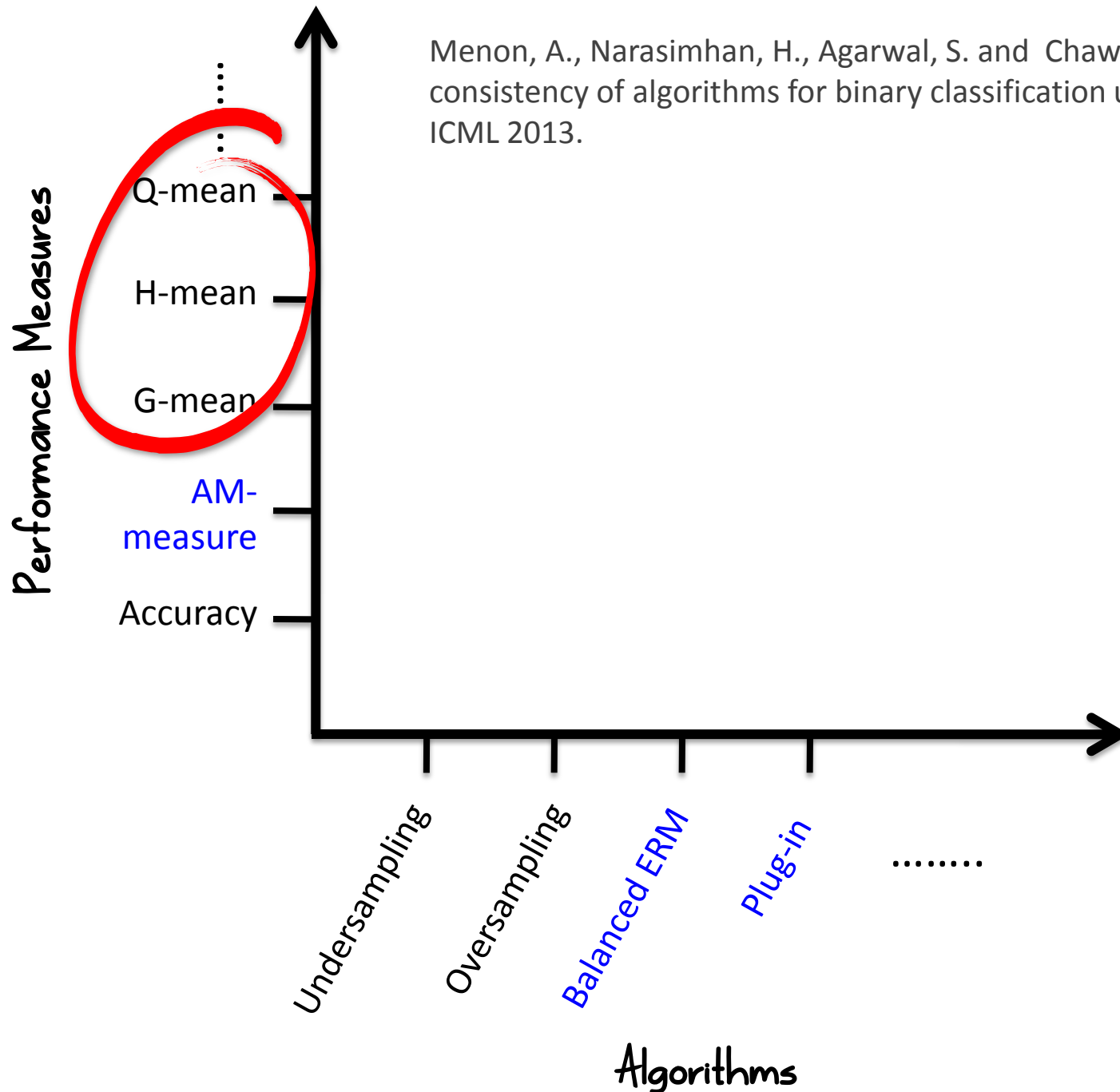


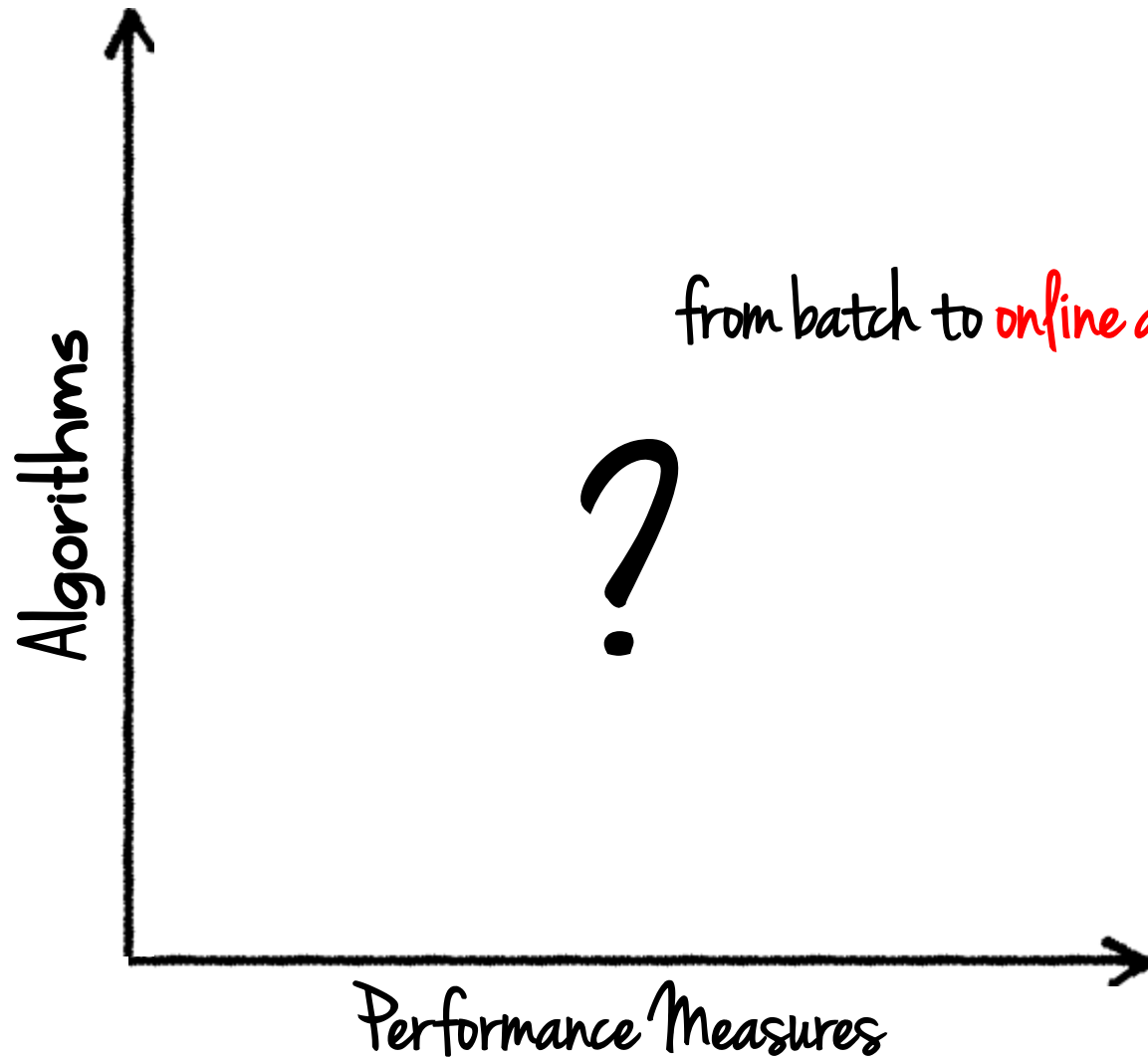
Standard Classification Error ||-suited!

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.





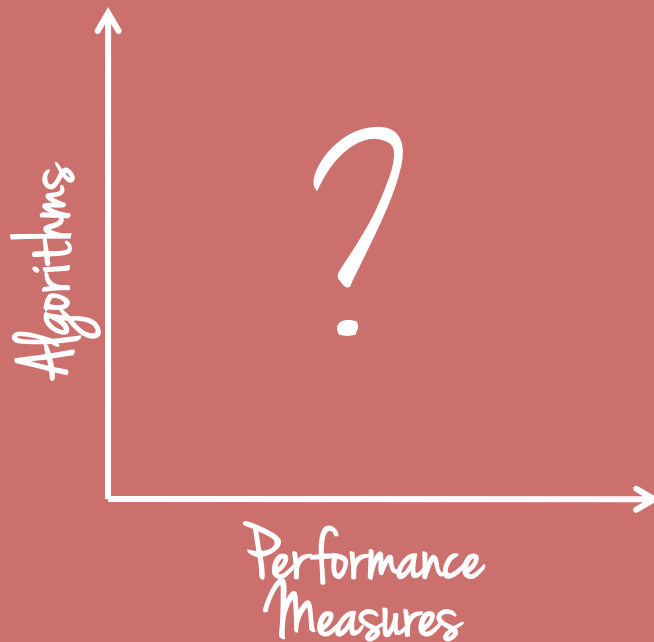
from batch to online algorithms ...

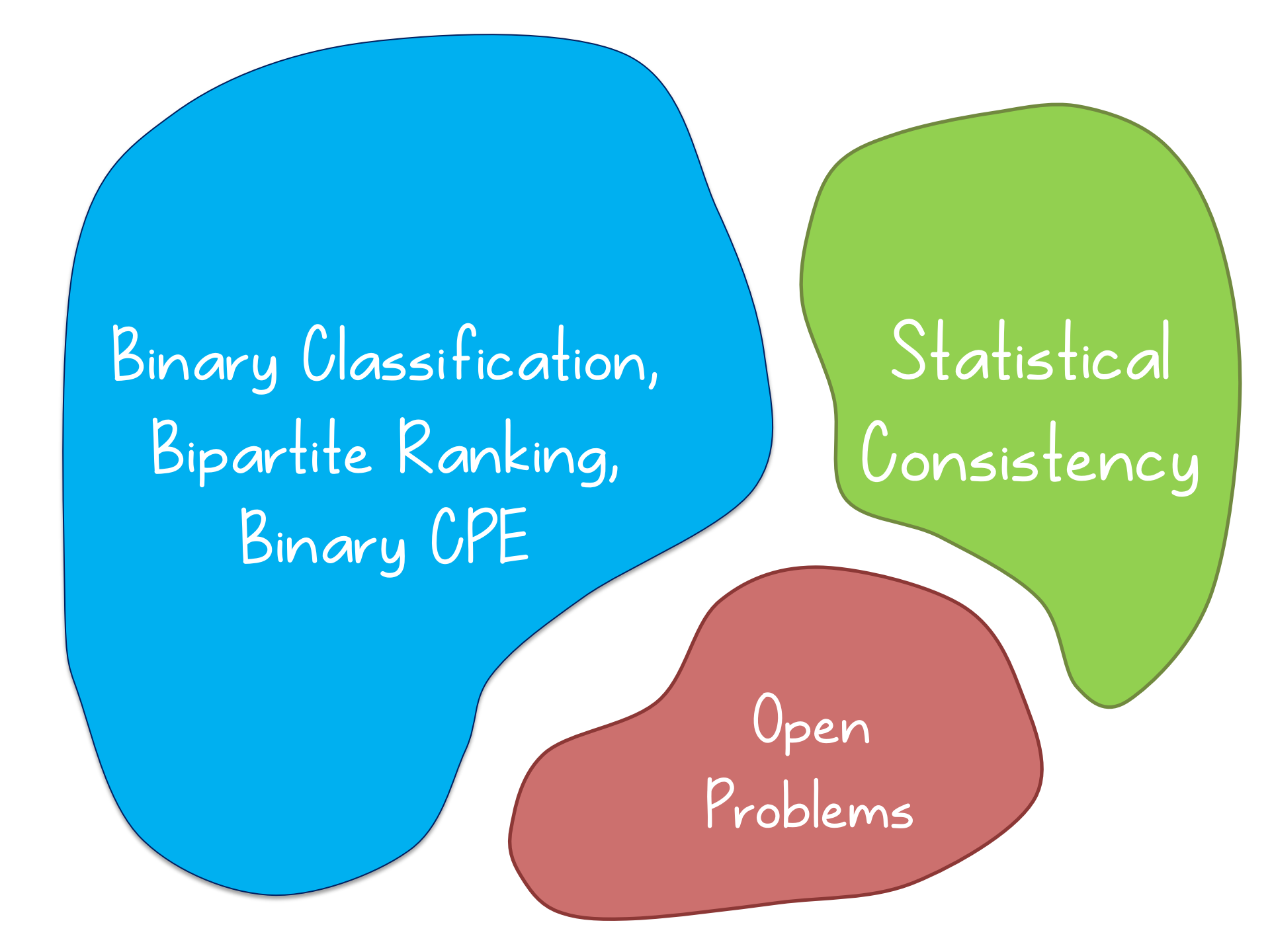
Batch Algorithms for Multivariate Performance Measures (Structural SVM)

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

Open Problems





Binary Classification,
Bipartite Ranking,
Binary CPE

Statistical
Consistency

Open
Problems

Questions?