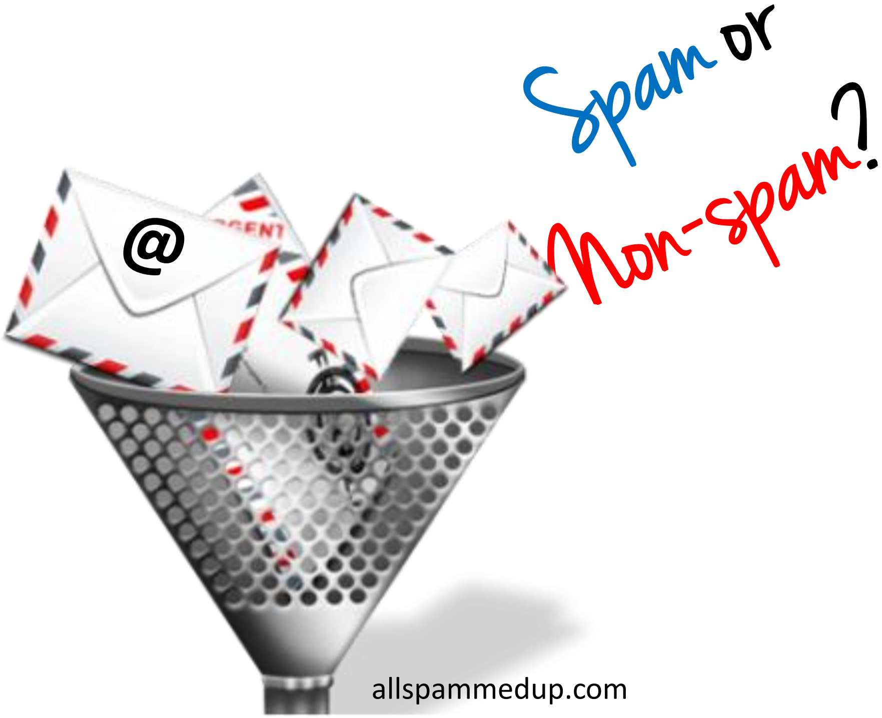
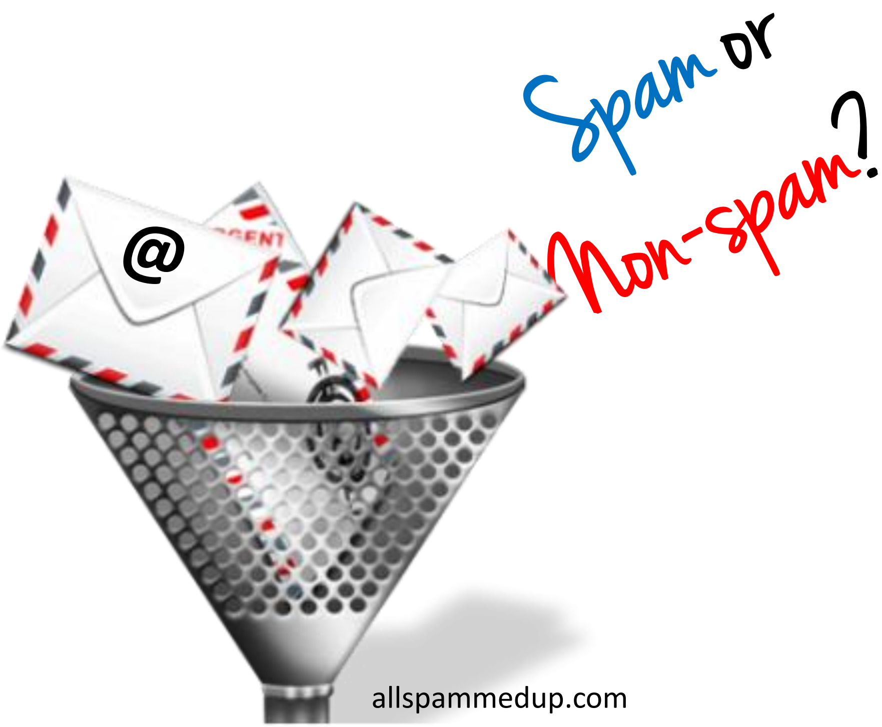
On the Relationship Between Binary Classification, Bipartite Ranking, and Binary Class Probability Estimation

Harikrishna Narasimhan

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


Spam or
Non-spam?
@
allspammedup.com


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

Relevant or Irrelevant?

fusionsedge.com

Spam or Non-spam?

allspammedup.com

Disease or No Disease?

pascal-network.org

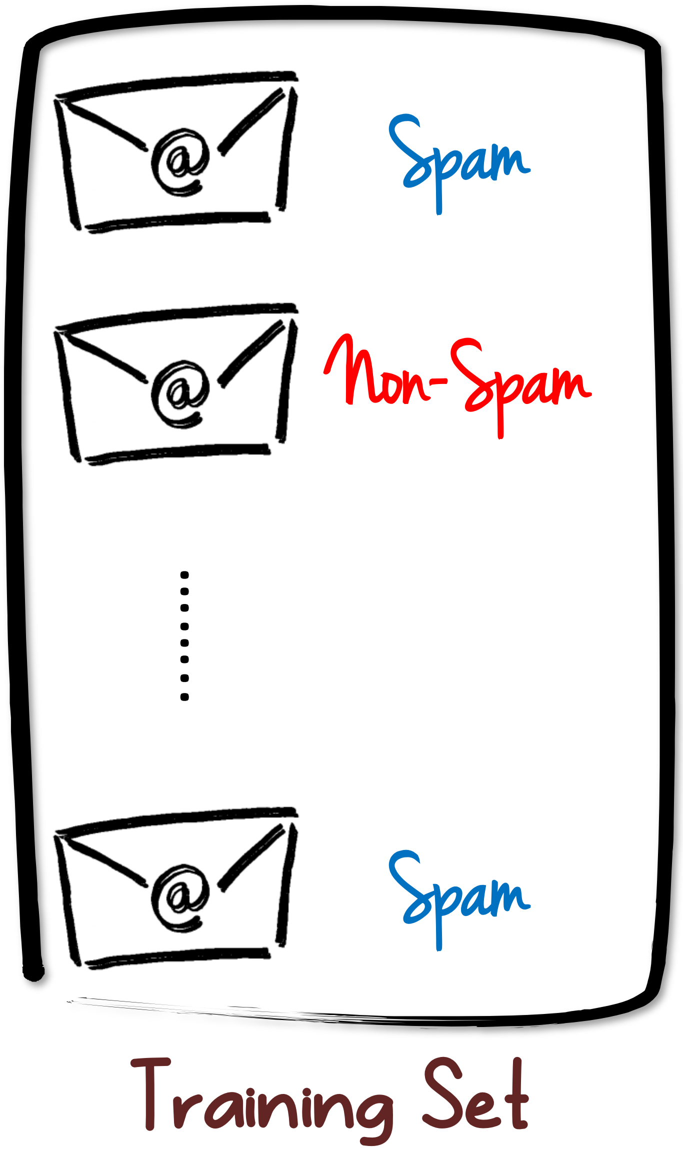
Relevant or Irrelevant?

fusionsedge.com

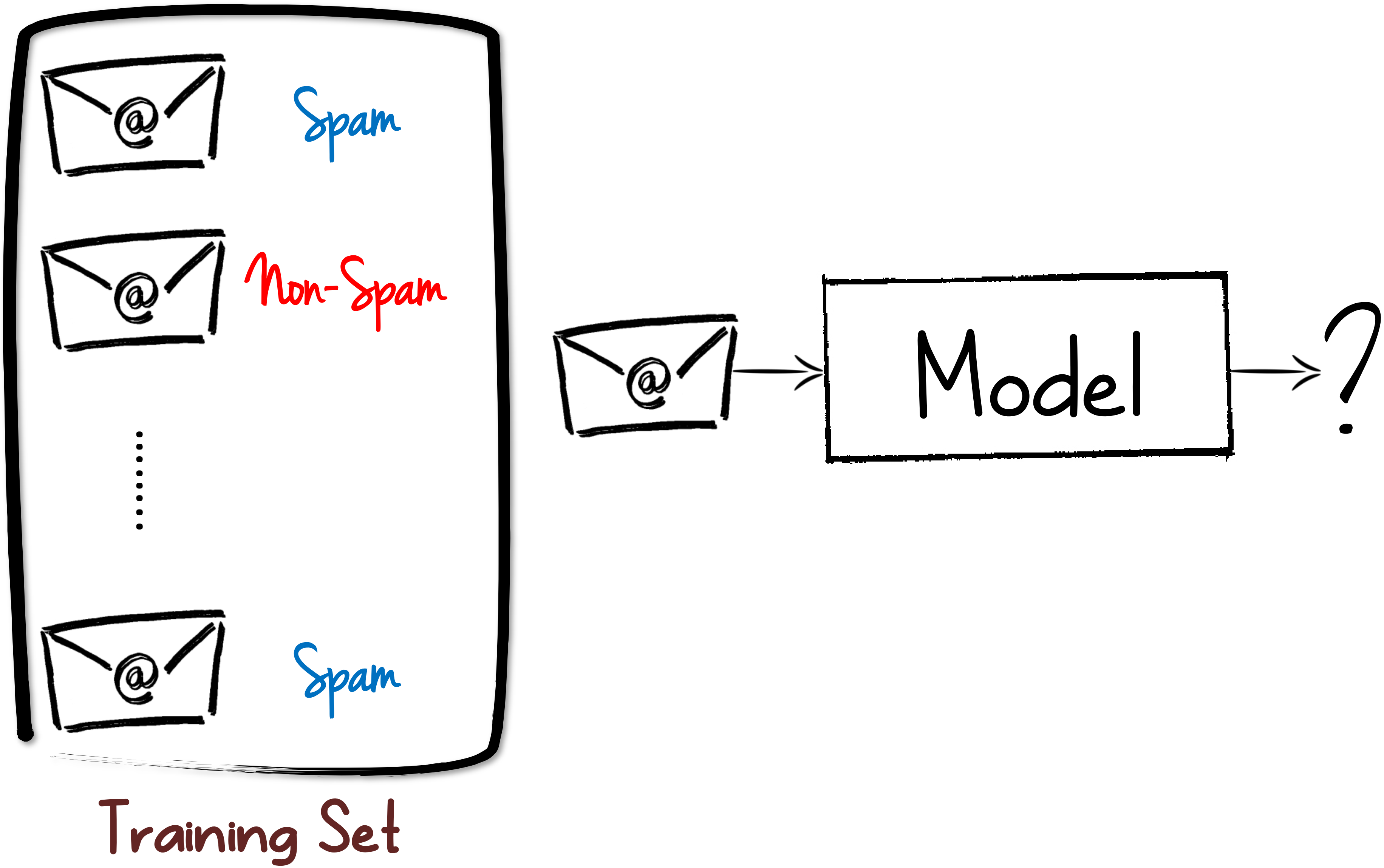
Click or No Click?

optimum7.com

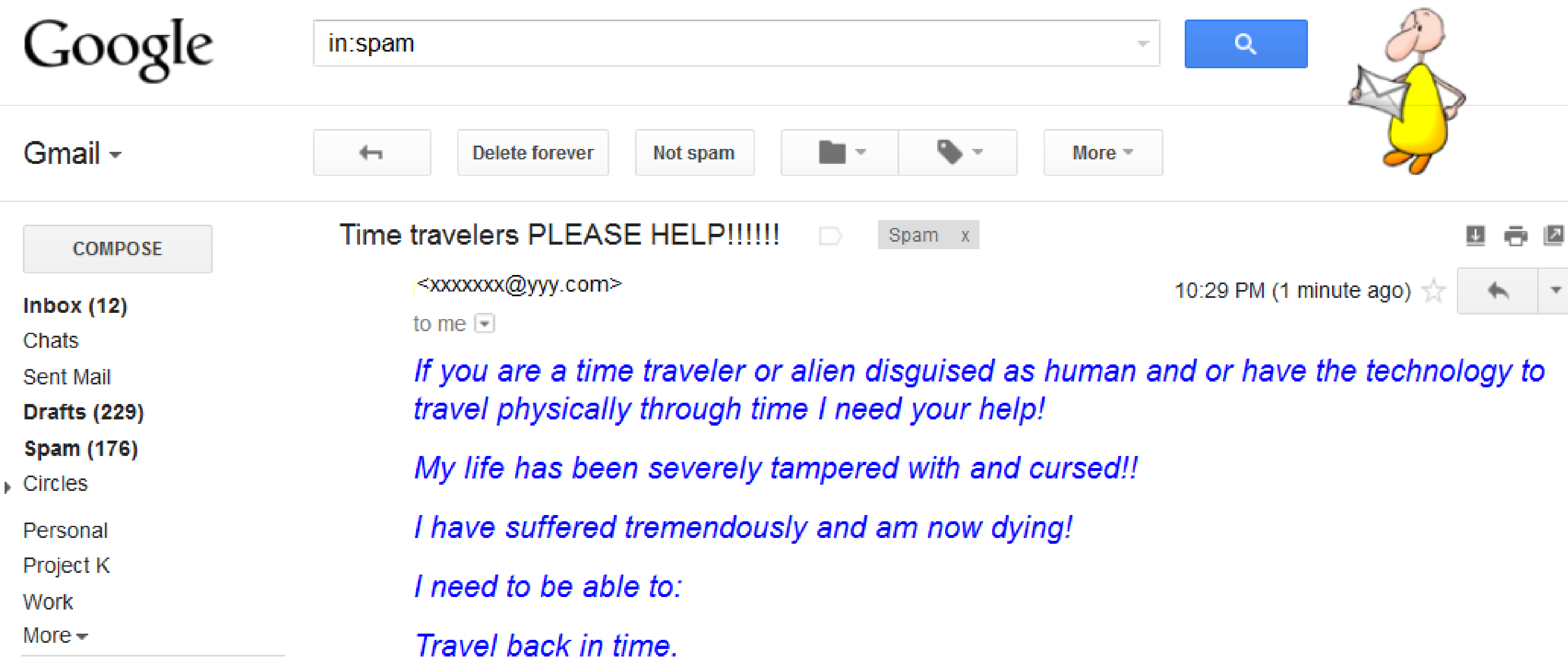
Learning from 'Binary' Labels



Learning from 'Binary' Labels



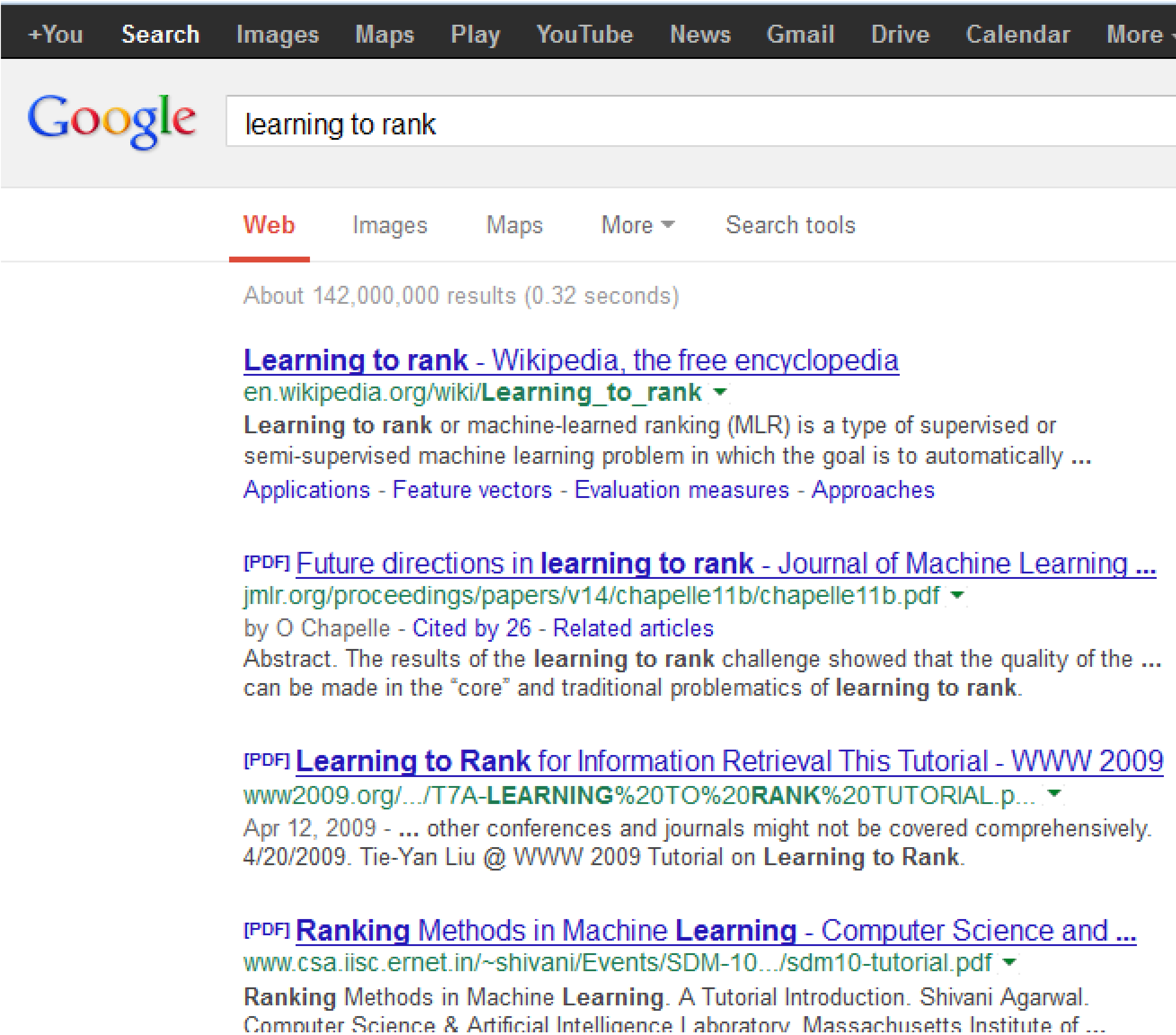
Binary Classification



Spam or Non-spam?

http://gmail.com/

Bipartite Ranking



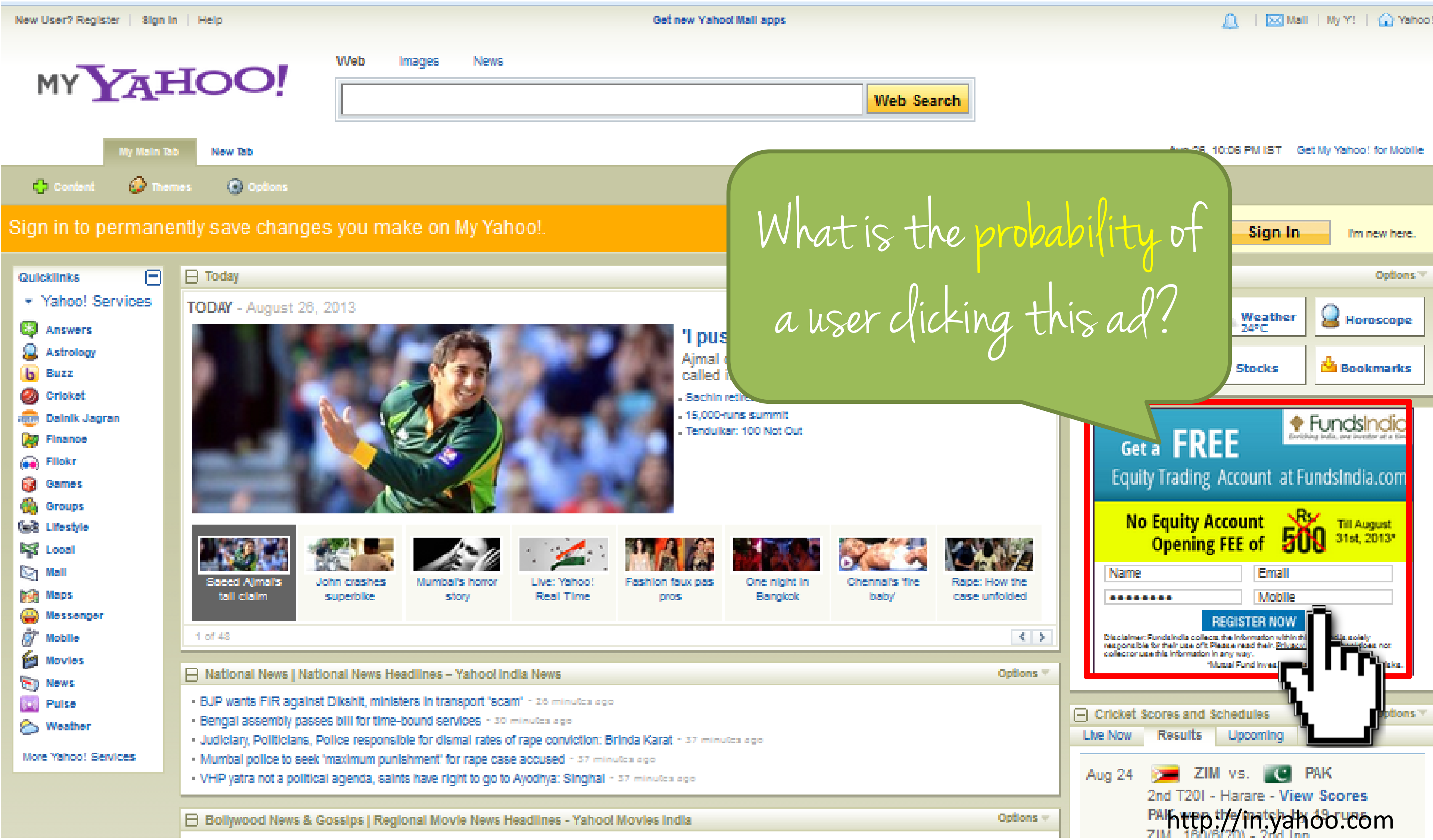
Relevant documents

above

Irrelevant documents

http://www.google.com/

Binary Class Probability Estimation



What is the probability of a user clicking this ad?

http://in.yahoo.com

three fundamental problems ...

Binary Classification

Bipartite Ranking

Binary Class Prob. Estimation

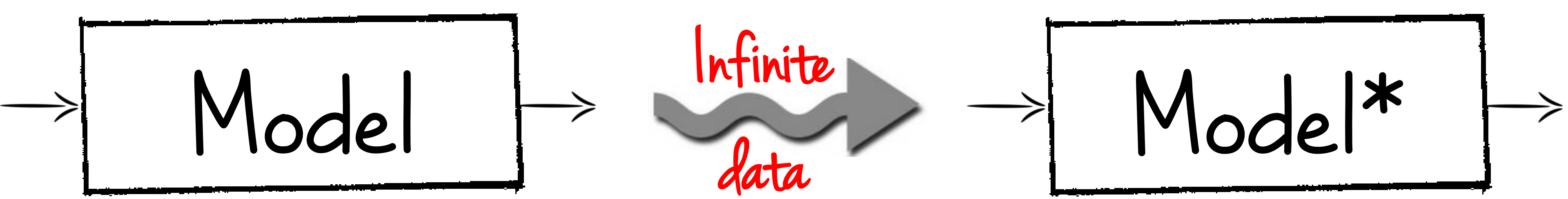
three fundamental problems ...

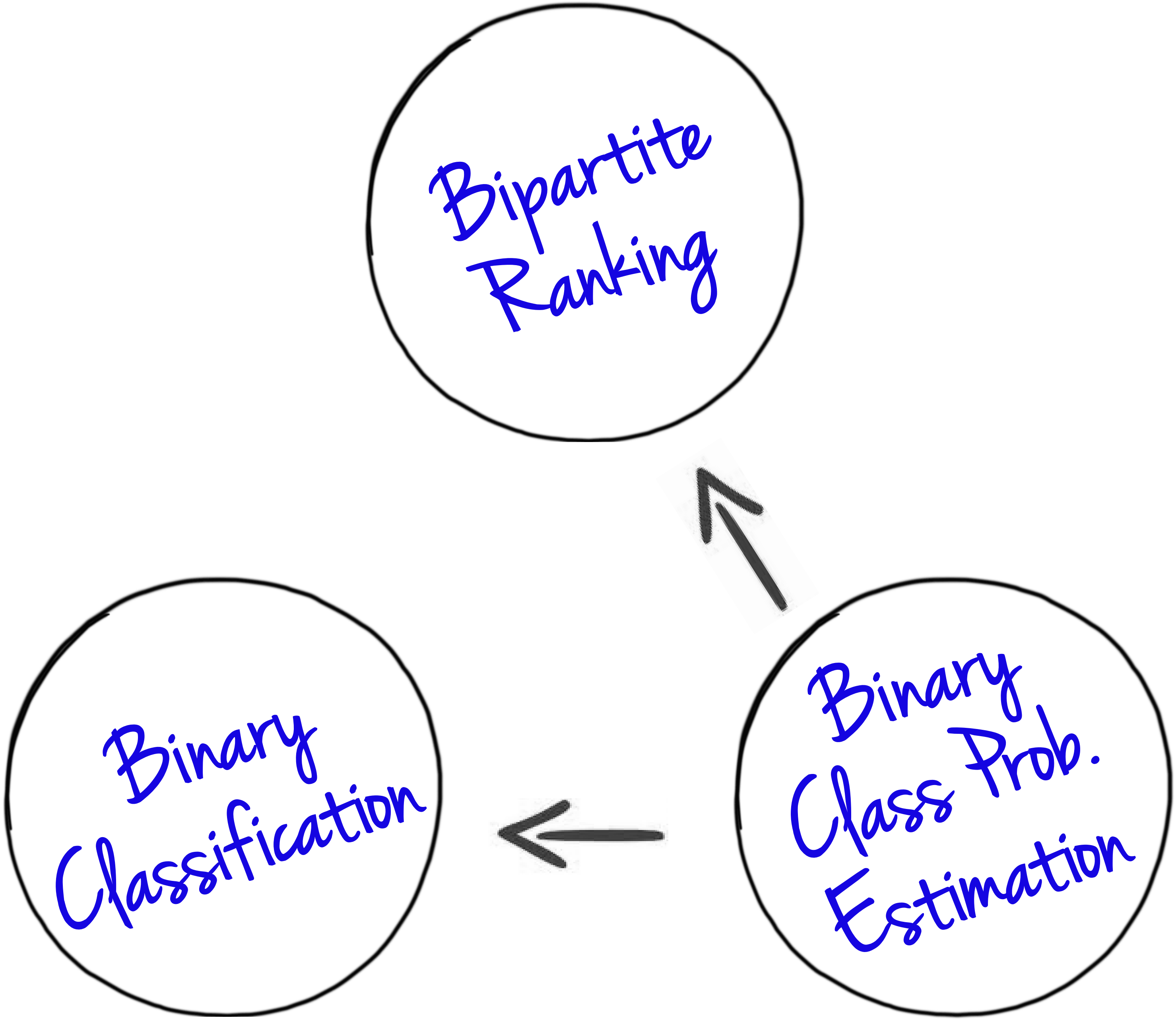
Binary Classification

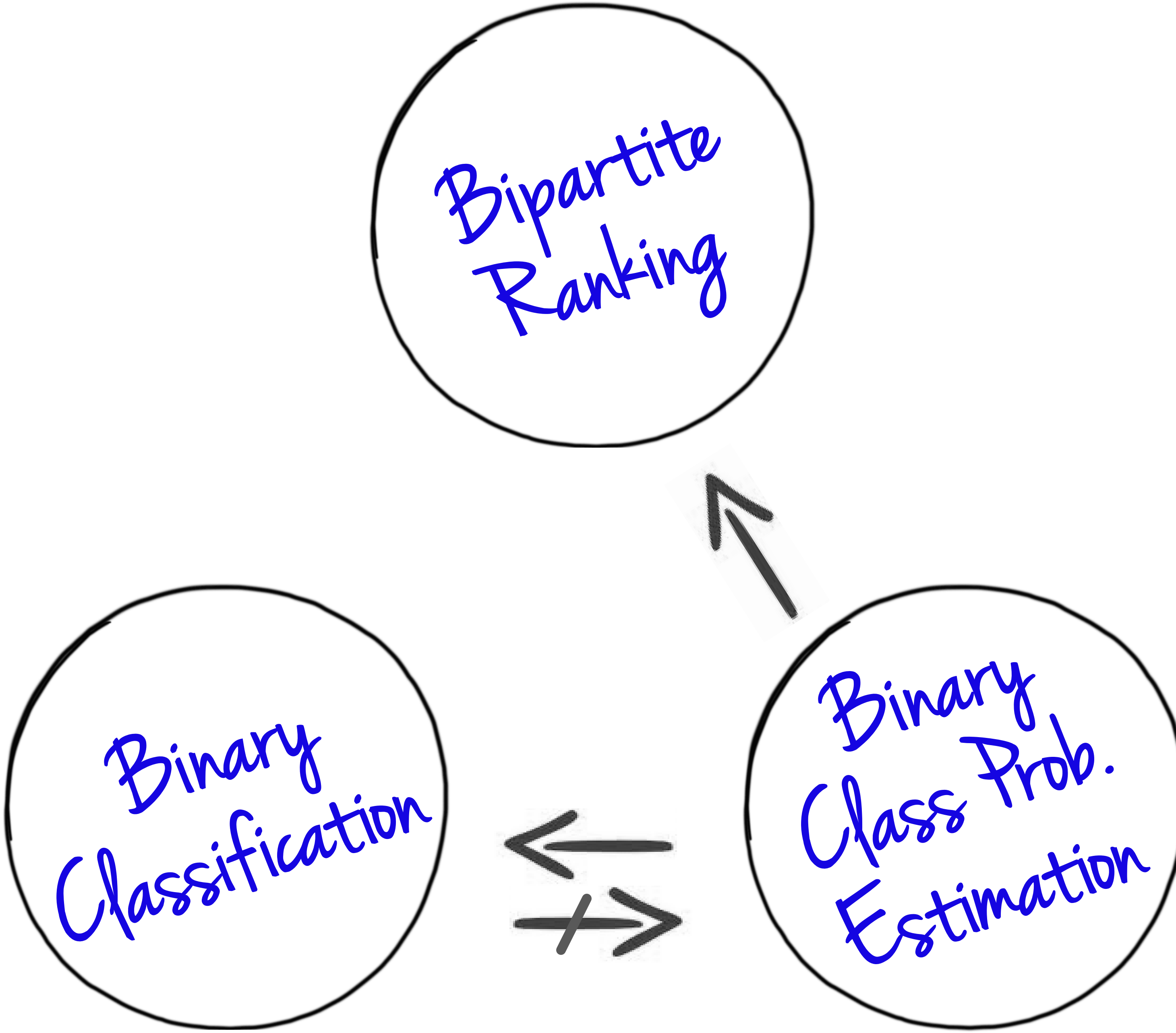
Bipartite Ranking

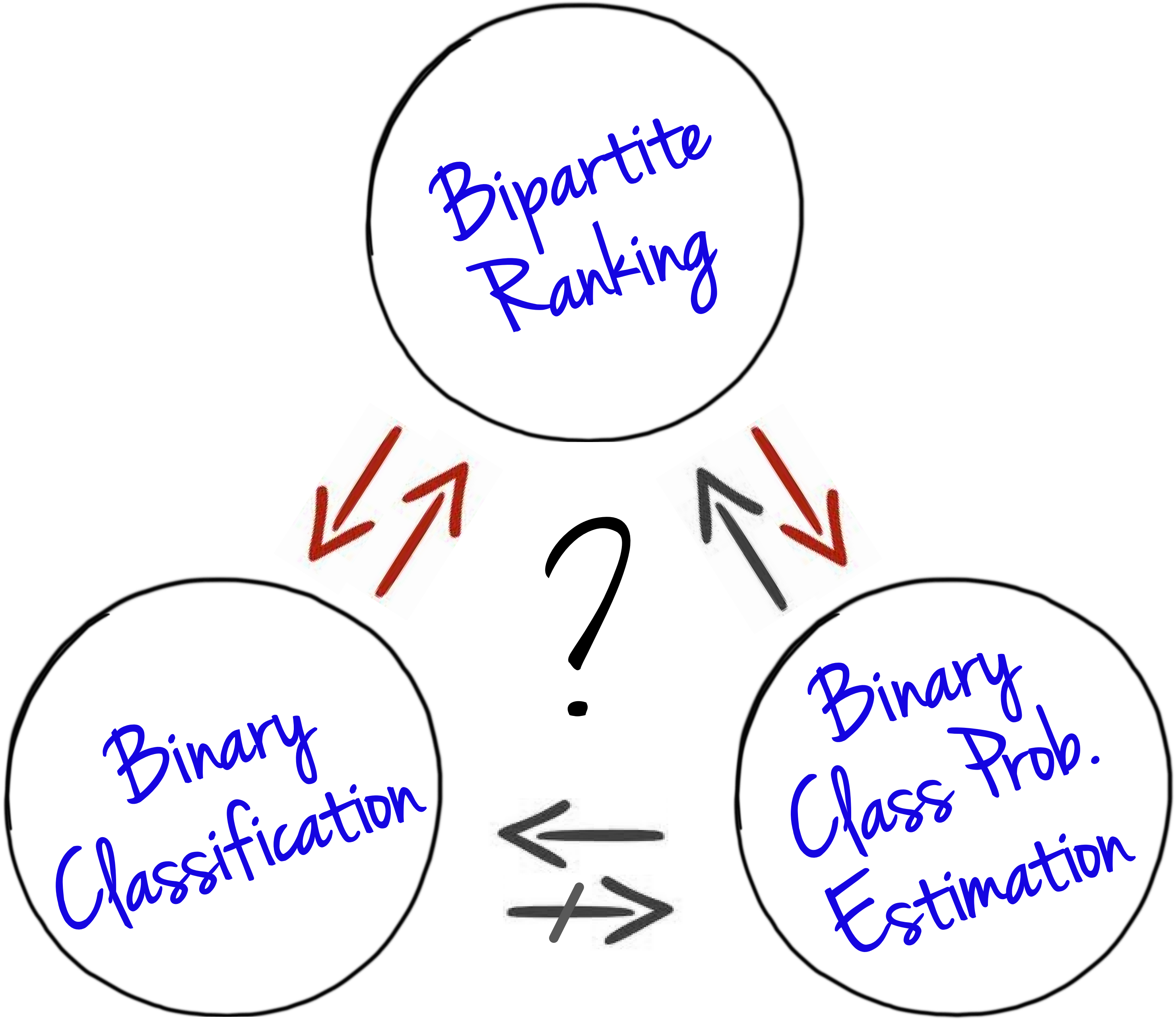
Binary Class Prob. Estimation

Statistical Consistency?

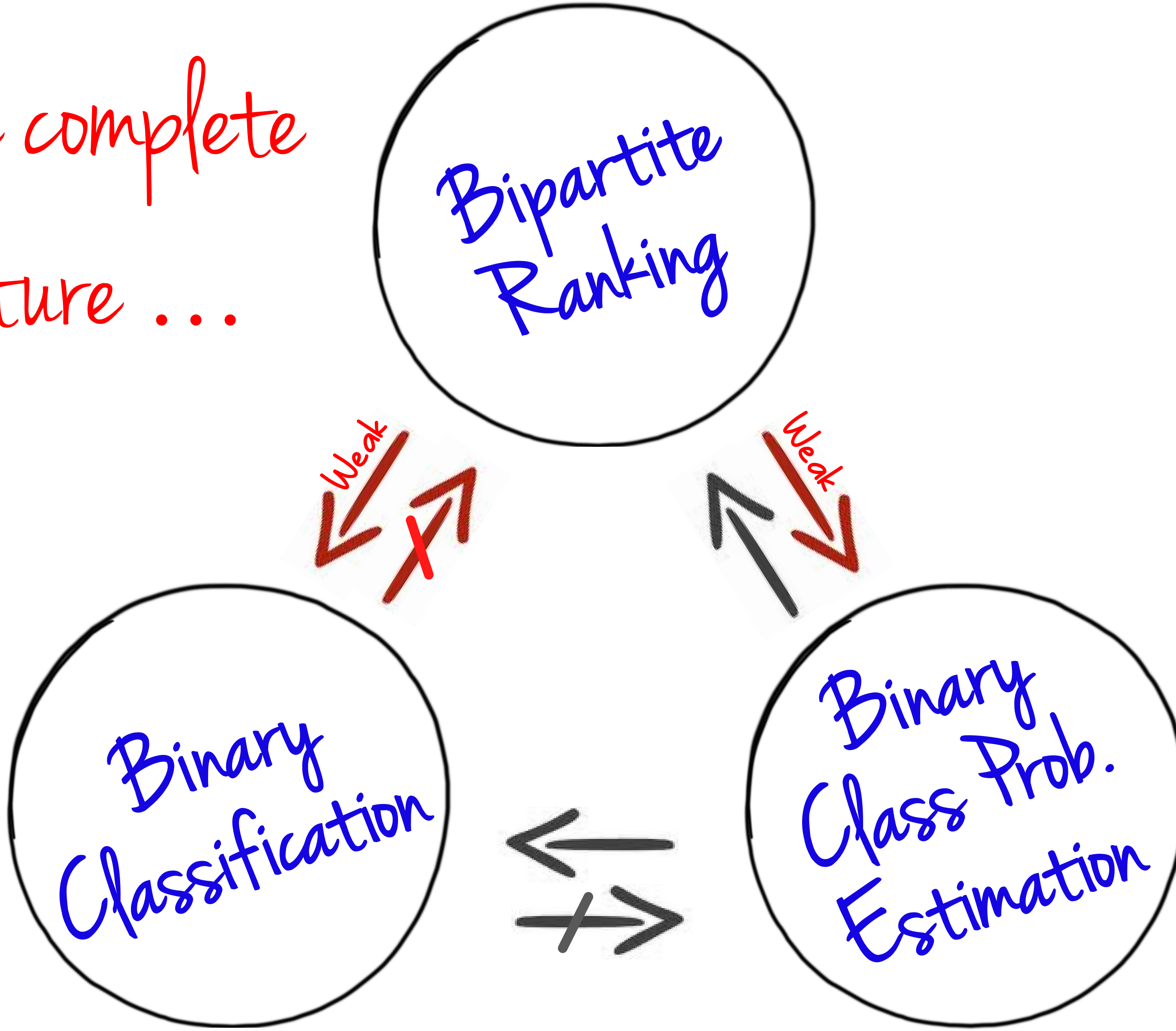



Bipartite Ranking
Binary Classification
Binary Class Prob. Estimation

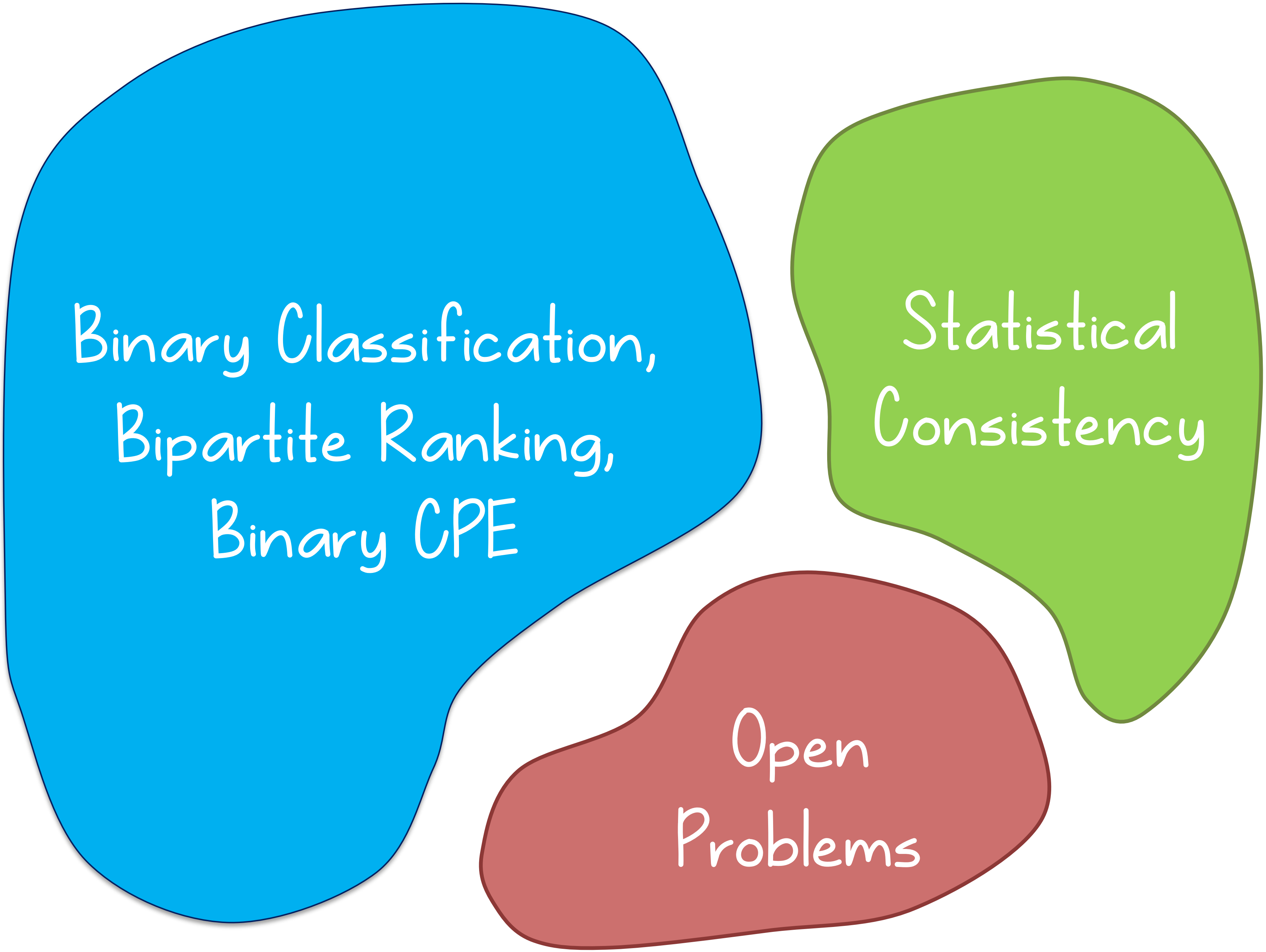

Bipartite Ranking
Binary Classification
Binary Class Prob. Estimation

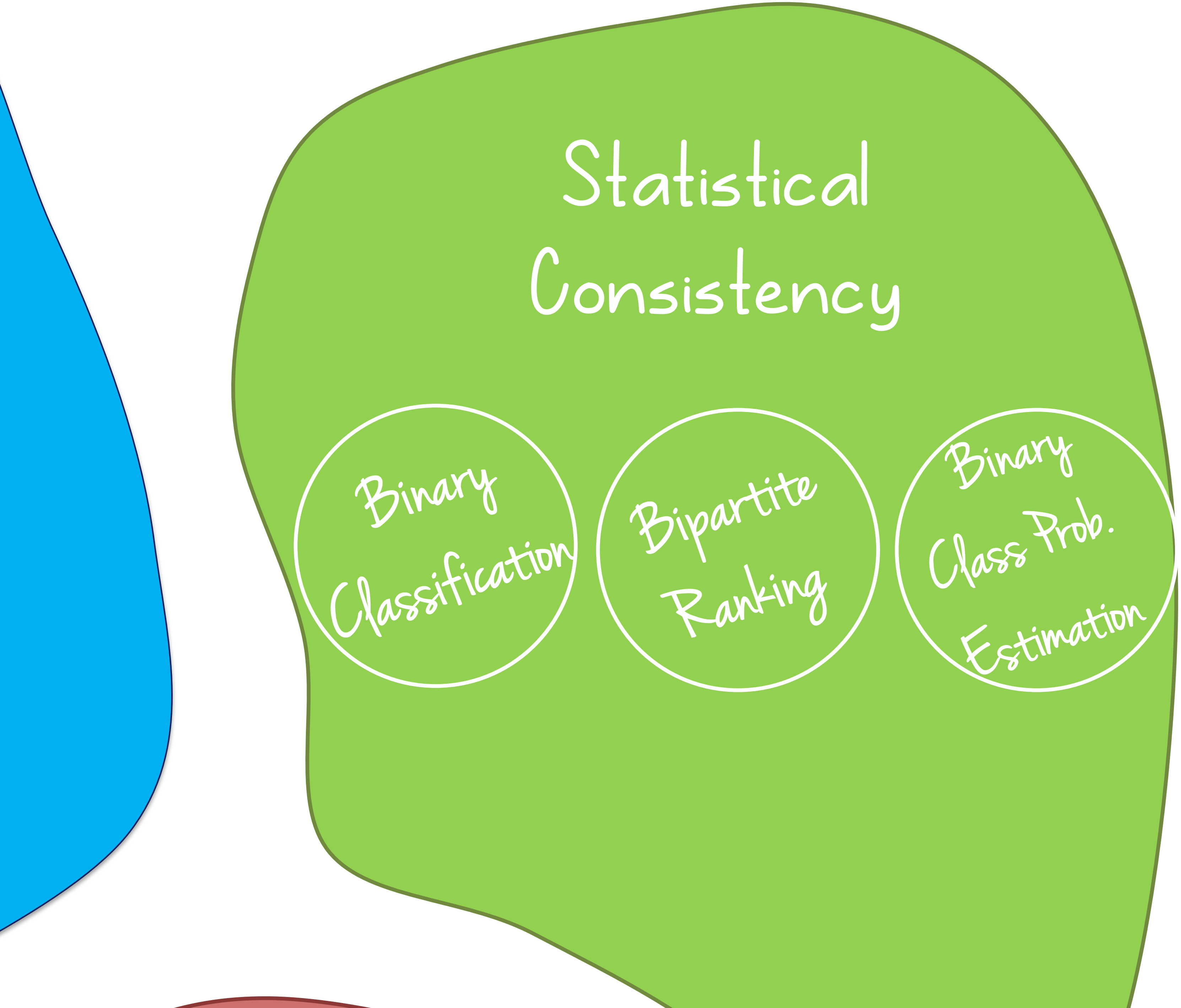

Bipartite Ranking
Binary Classification
Binary Class Prob. Estimation
?

the complete picture ...



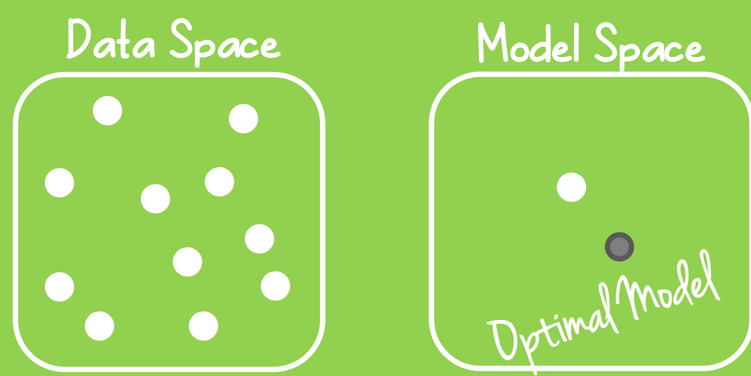
Narasimhan, H. and Agarwal, S. "*On the relationship between binary classification, bipartite ranking, and binary class probability estimation*". In NIPS 2013. To appear.


Binary Classification,
Bipartite Ranking,
Binary CPE
Statistical
Consistency
Open
Problems

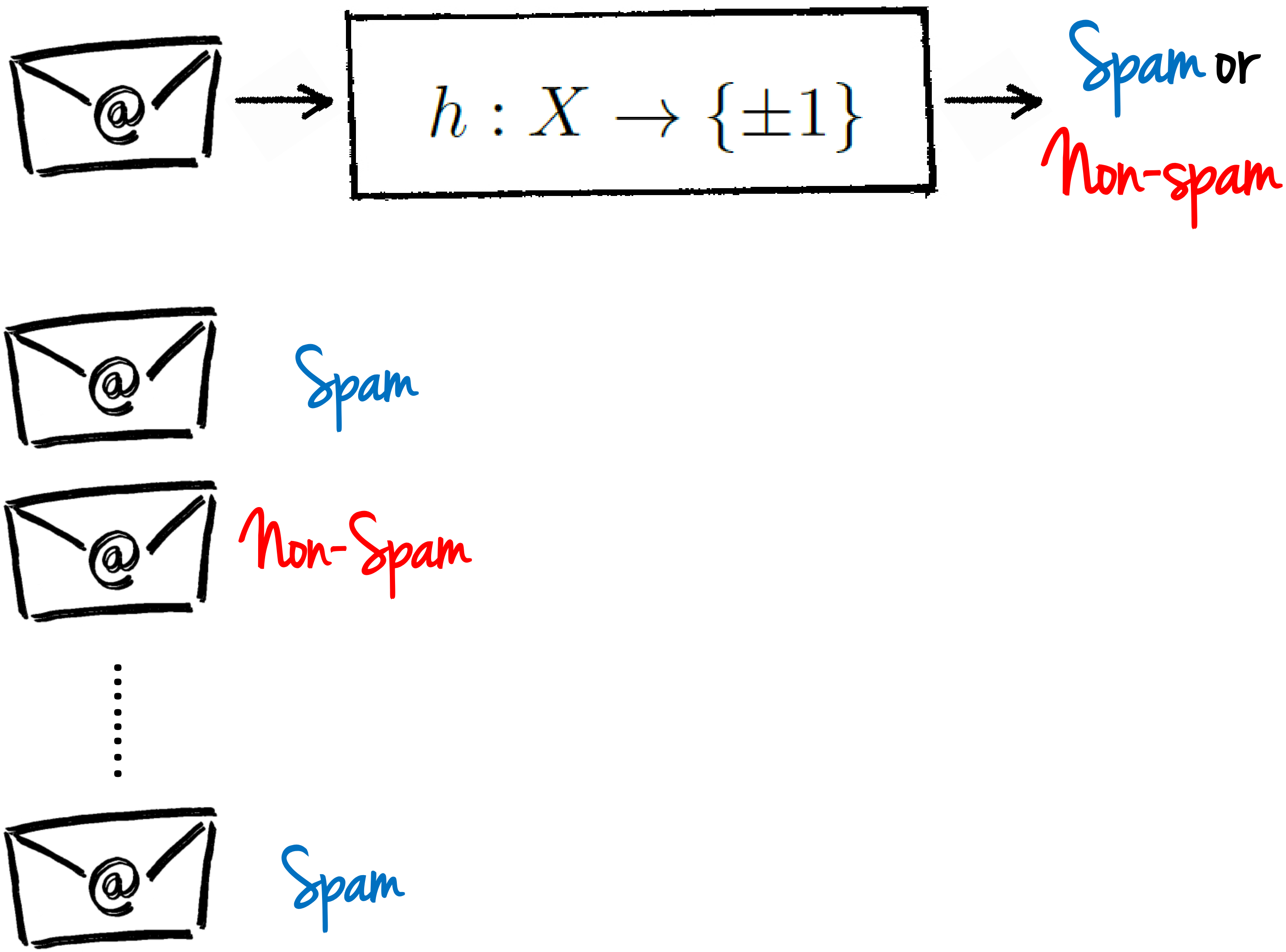


Statistical
Consistency
Binary
Classification
Bipartite
Ranking
Binary
Class Prob.
Estimation

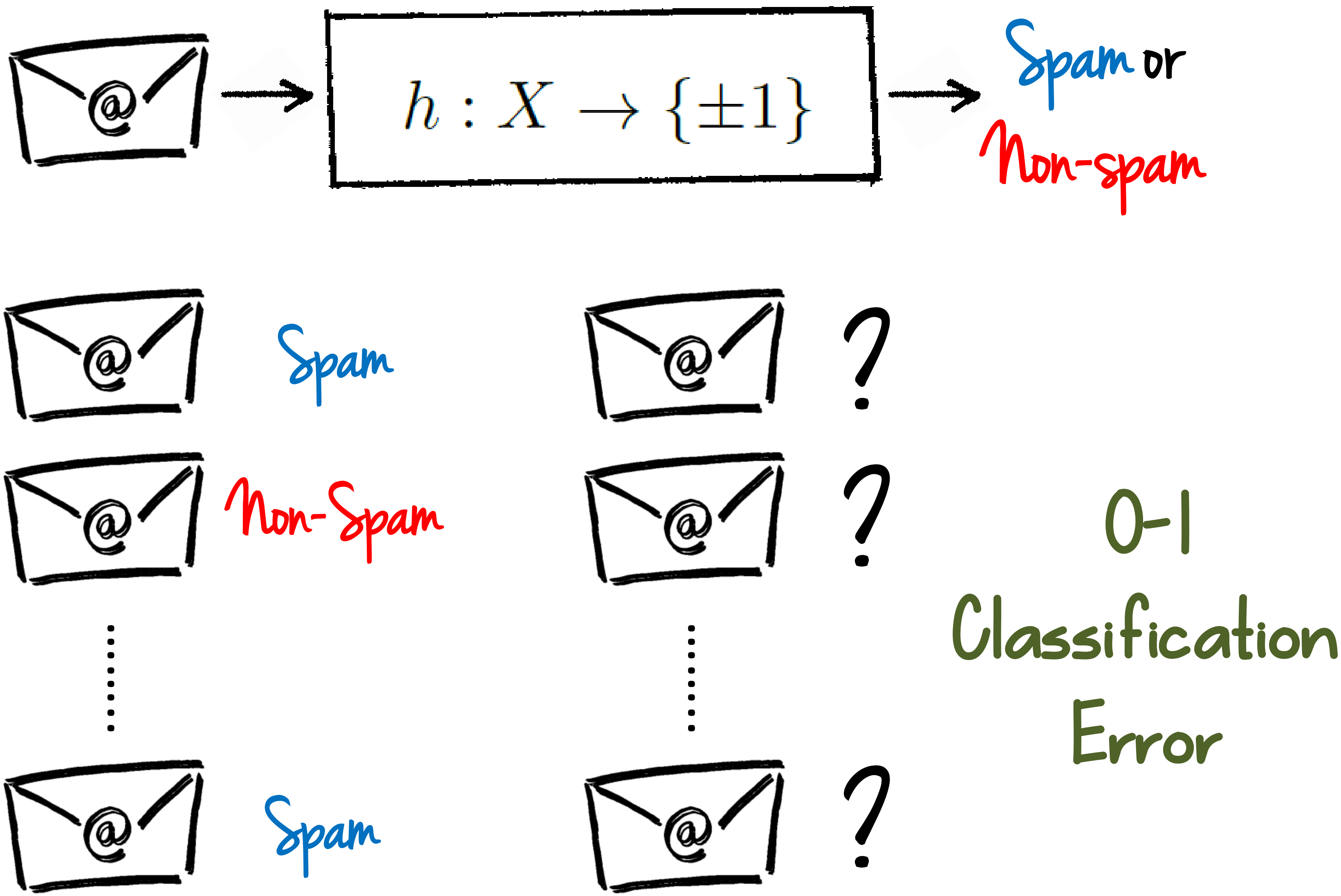
Statistical Consistency



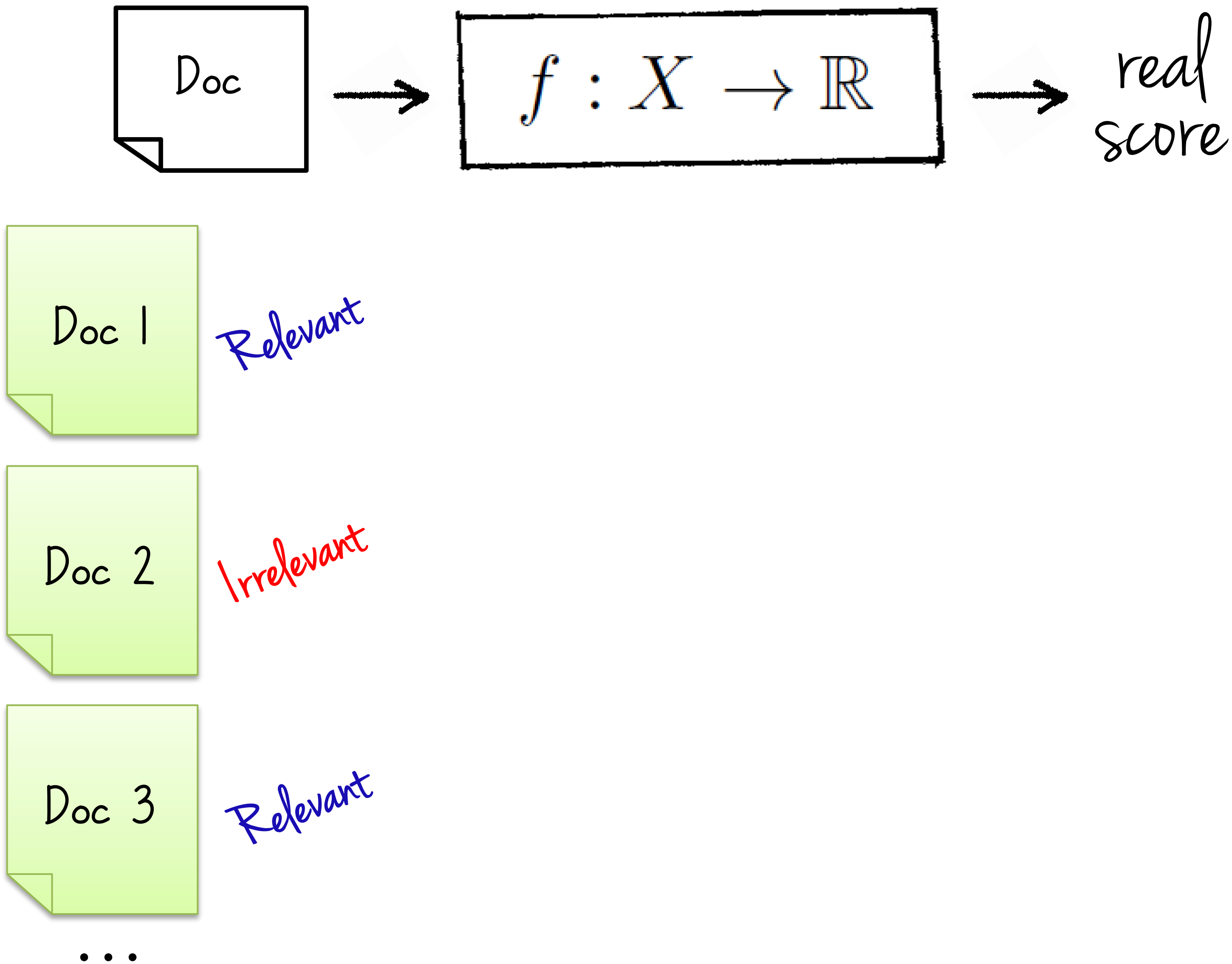
Binary Classification



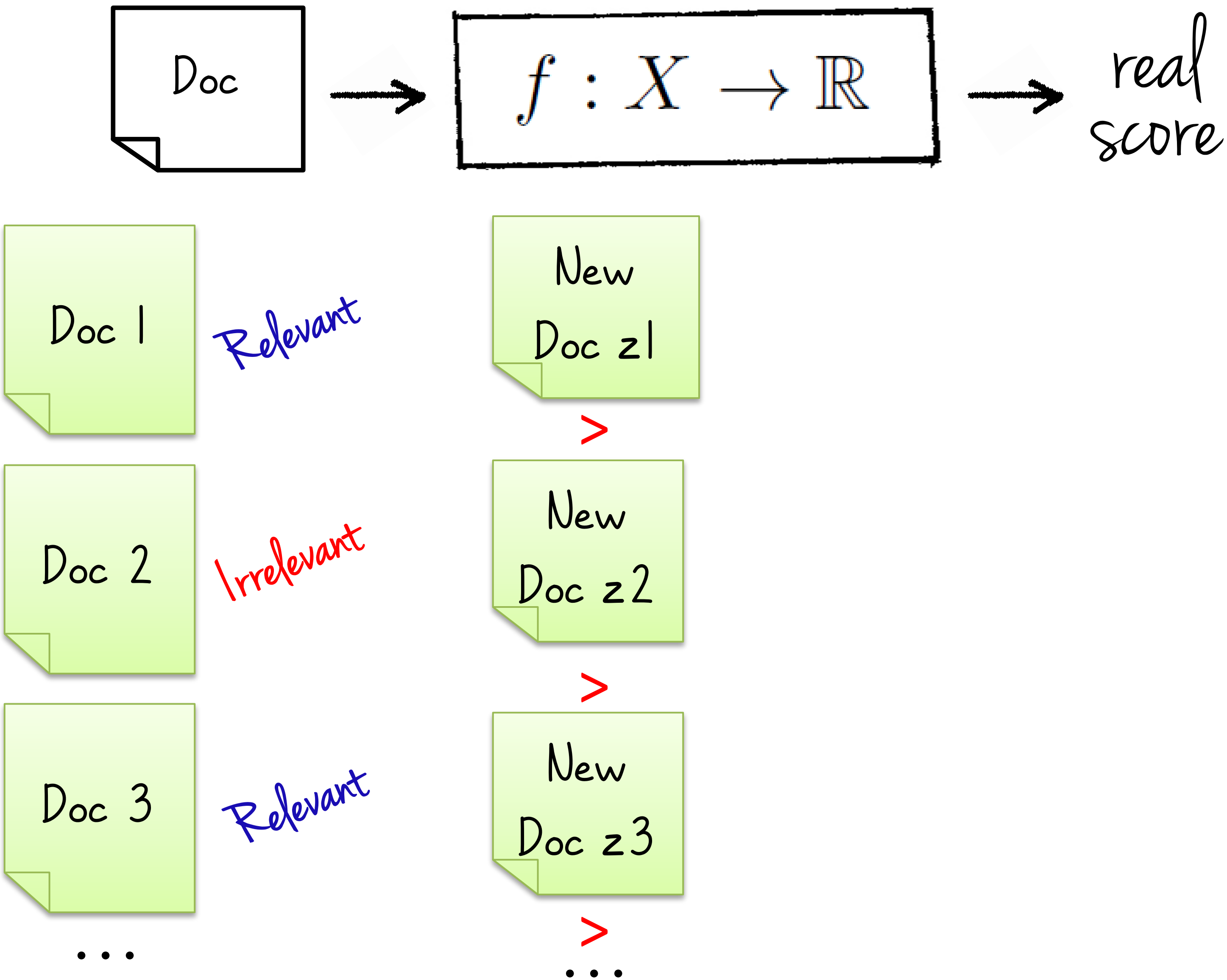
Binary Classification



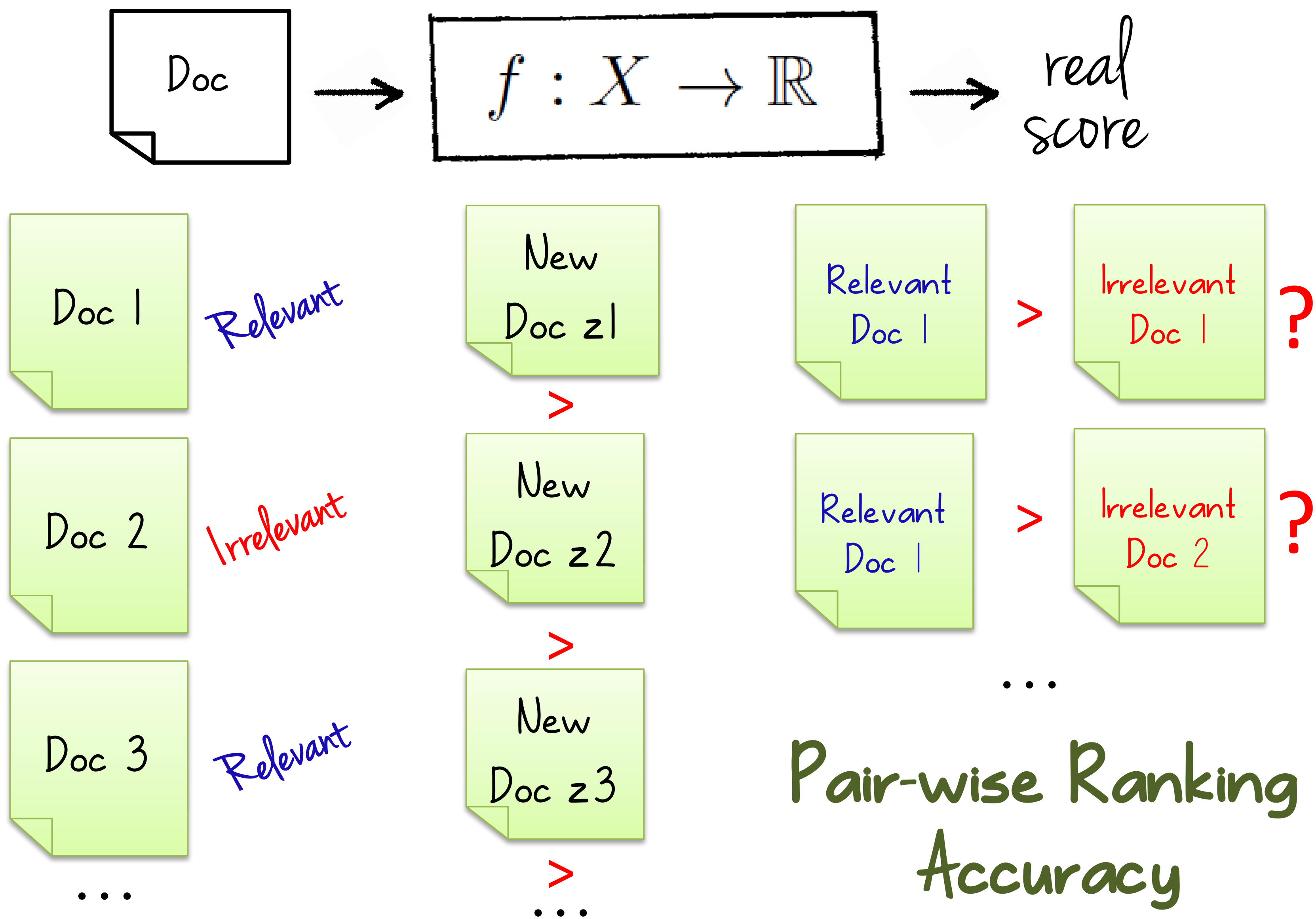
Bipartite Ranking



Bipartite Ranking



Bipartite Ranking



Binary Class Probability Estimation

Ad $\rightarrow$ $\boxed{\widehat{\eta} : X \rightarrow [0, 1]}$ $\rightarrow$ probability of click

Ad 1 Click

Ad 2 No Click

Ad 3 Click

...

Binary Class Probability Estimation

Ad → $\widehat{\eta} : X \rightarrow [0, 1]$ → probability of click

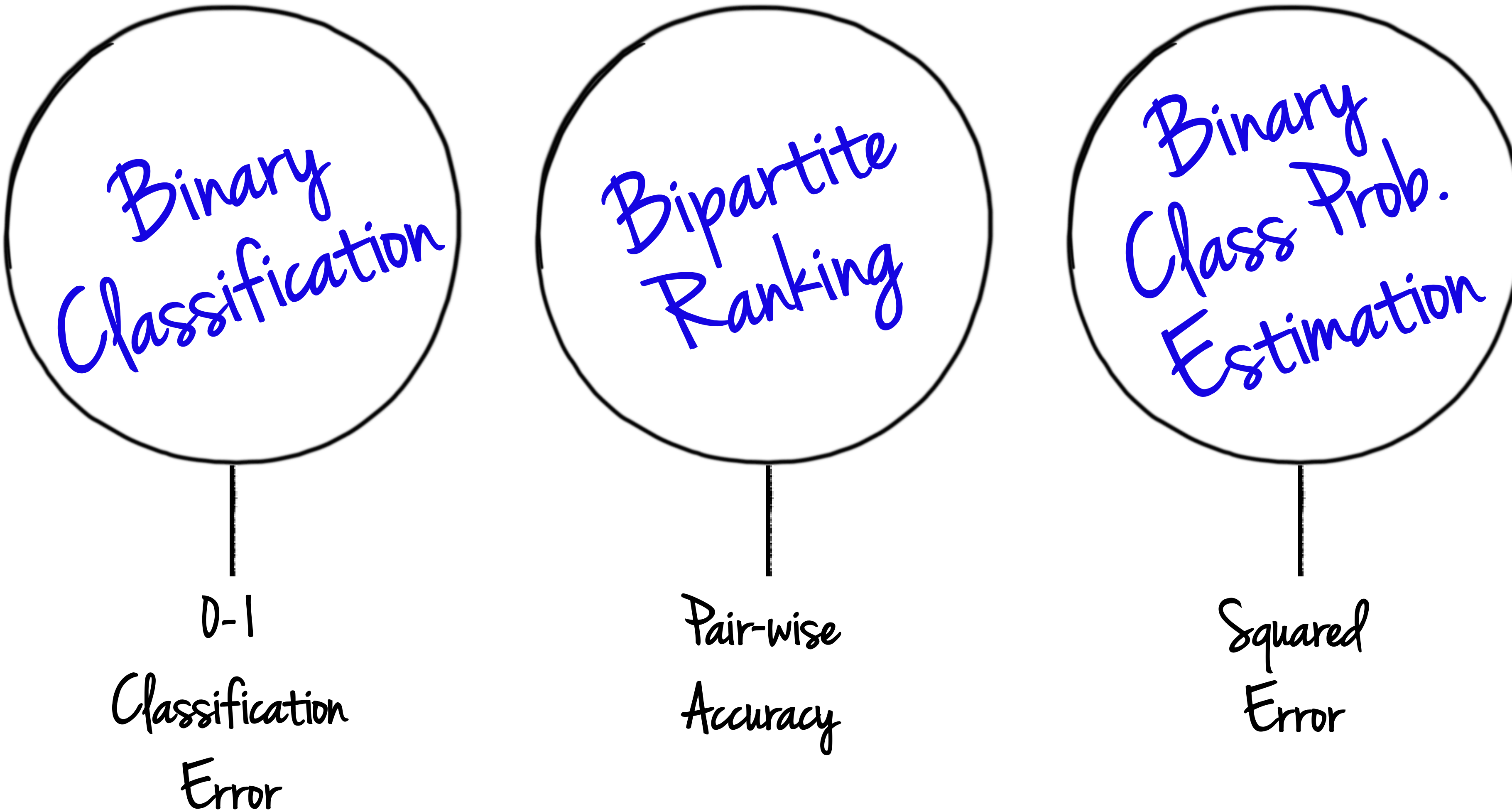
Ad 1	Click	New Ad 1	0.1
Ad 2	No Click	New Ad 2	0.4
Ad 3	Click	New Ad 3	0.2
...		...	

Binary Class Probability Estimation

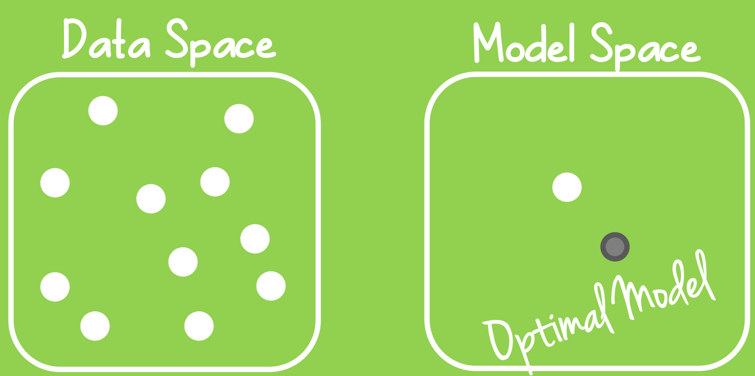
Ad → $\widehat{\eta} : X \to [0, 1]$ → probability of click

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
...		...		

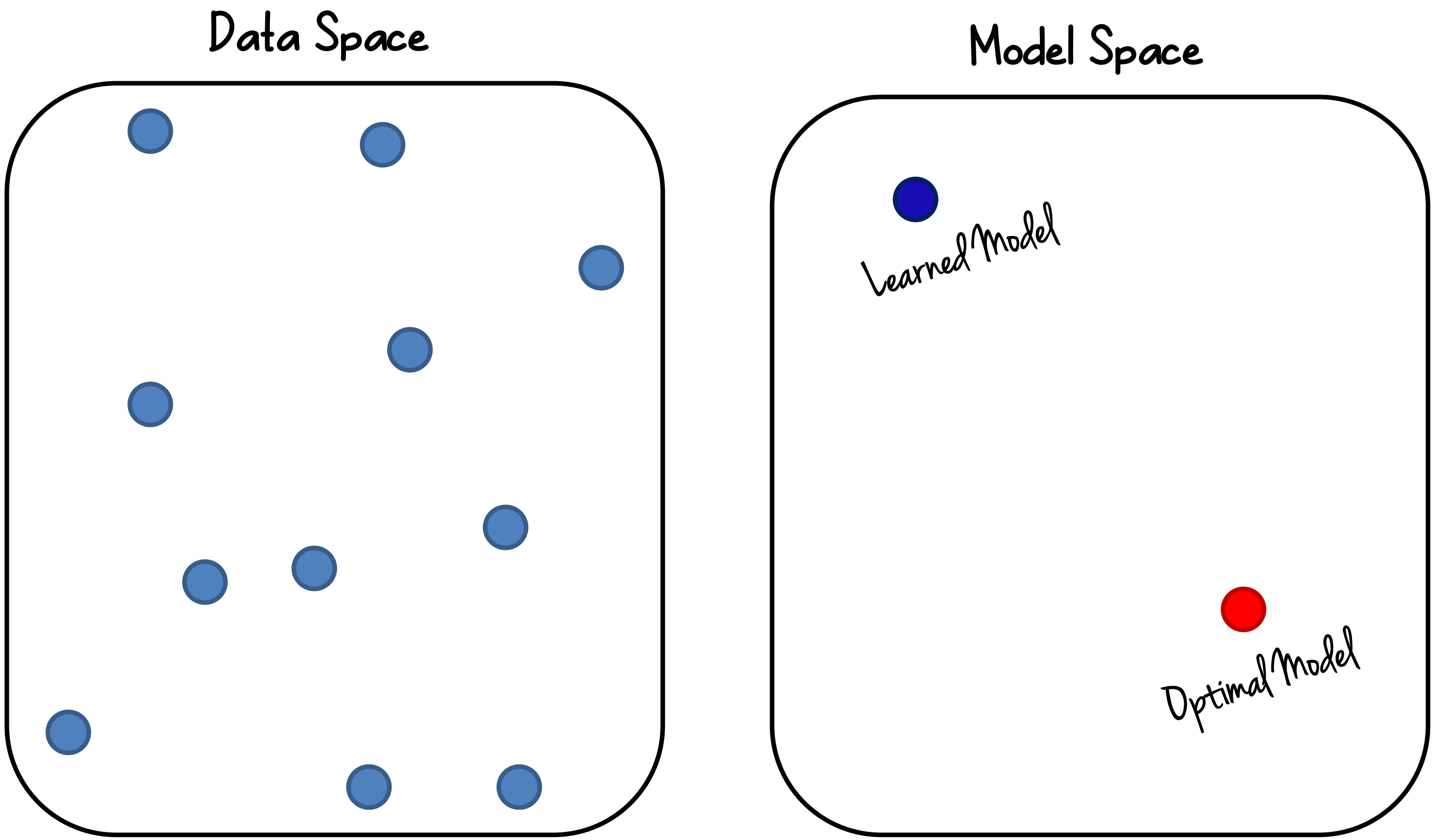
Squared Error


Binary Classification
Bipartite Ranking
Binary Class Prob. Estimation
0-1 Classification Error
Pair-wise Accuracy
Squared Error

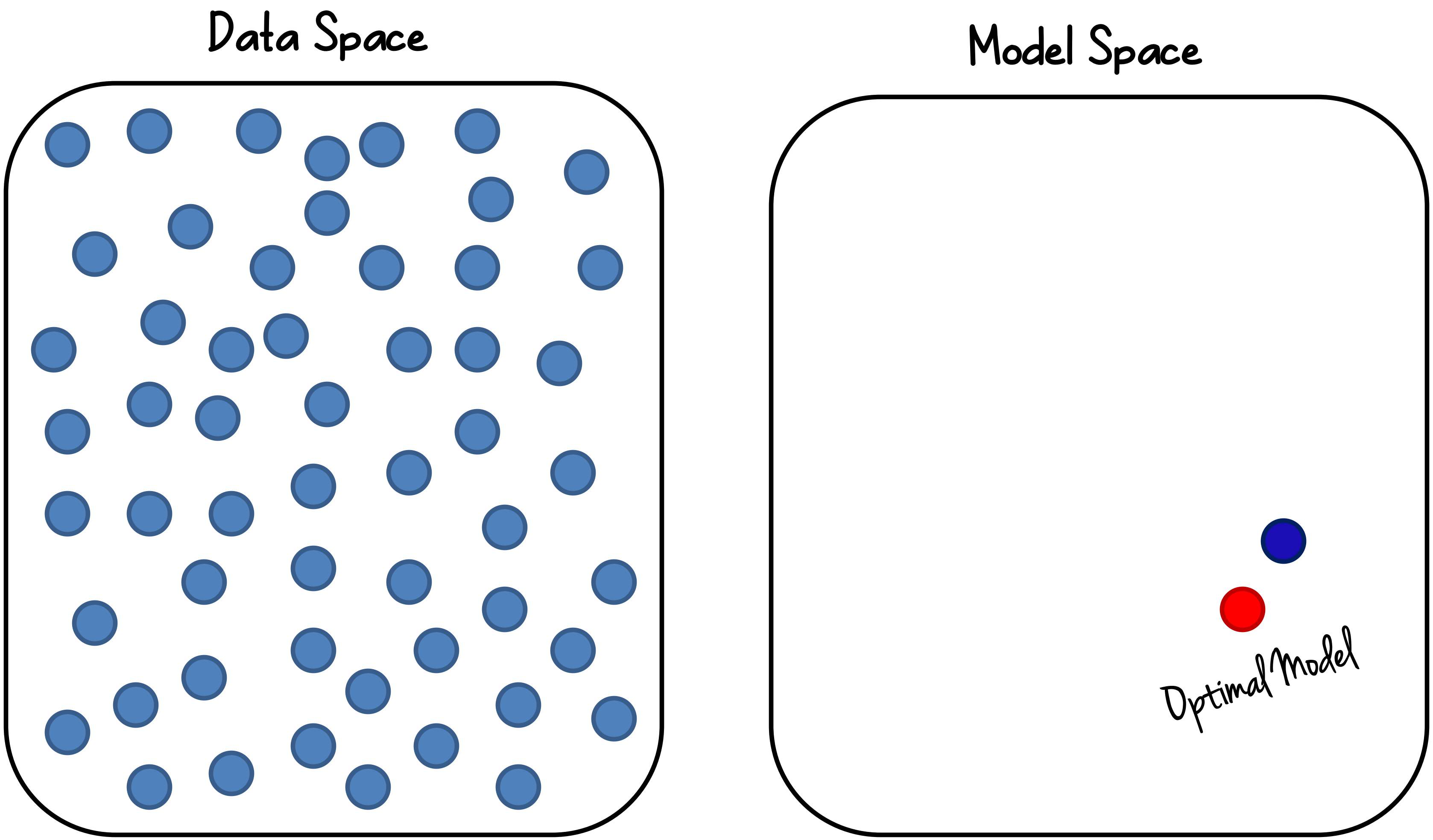
Statistical Consistency



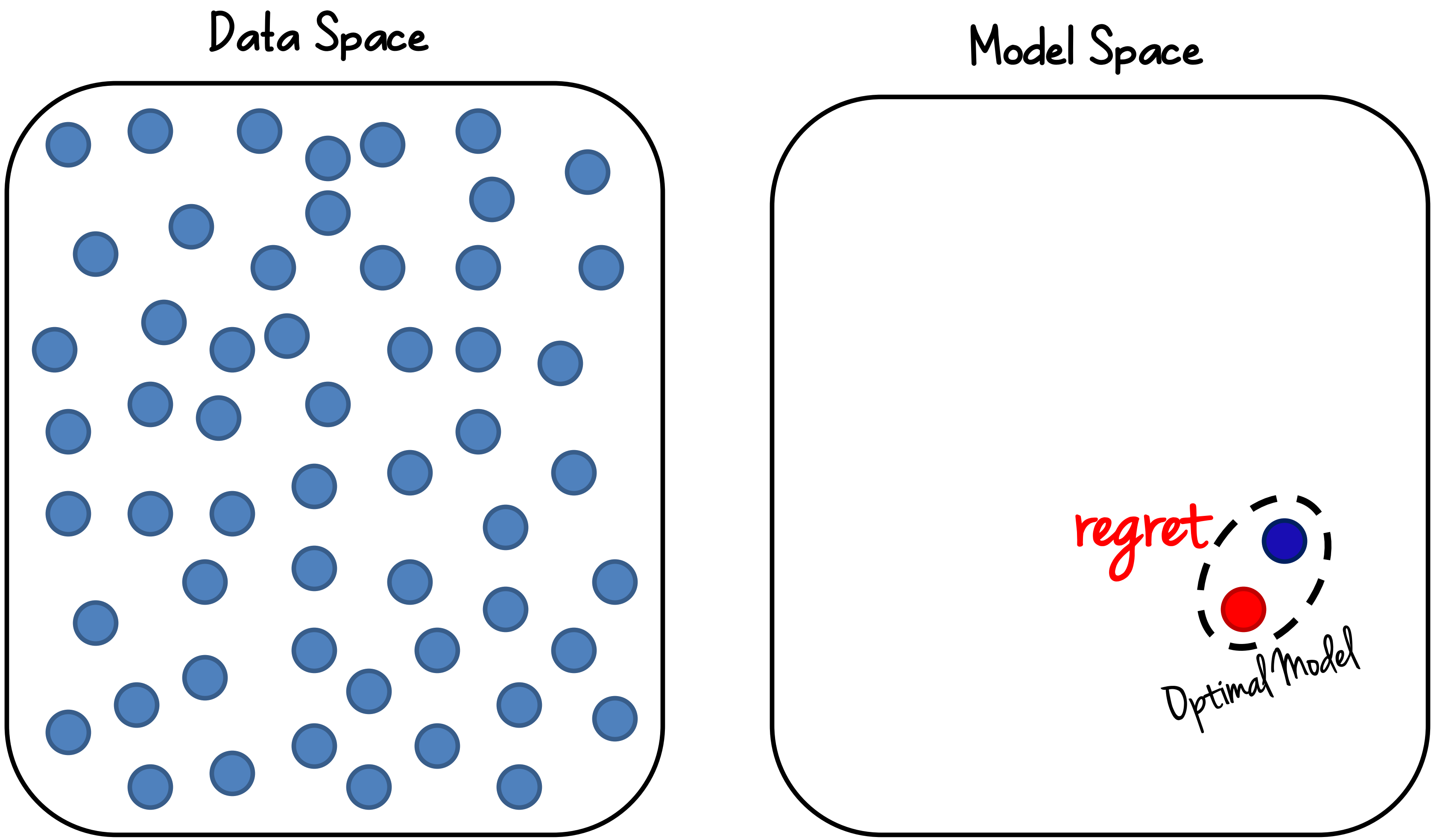
Statistical Consistency



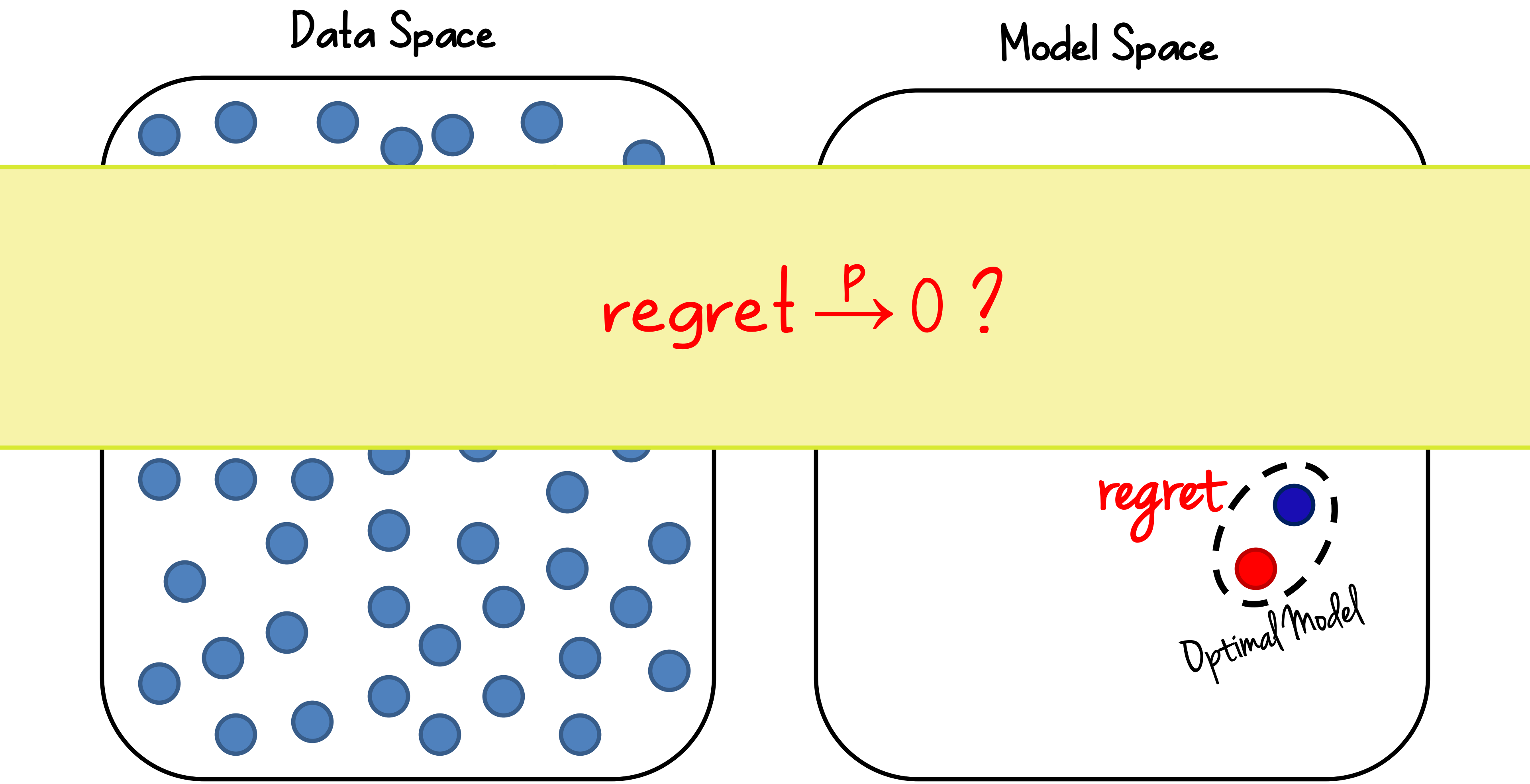
Statistical Consistency



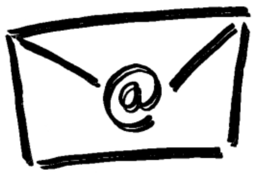
Statistical Consistency



Statistical Consistency



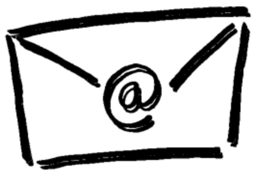
Binary Classification



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

Spam or Non-spam

Binary Classification

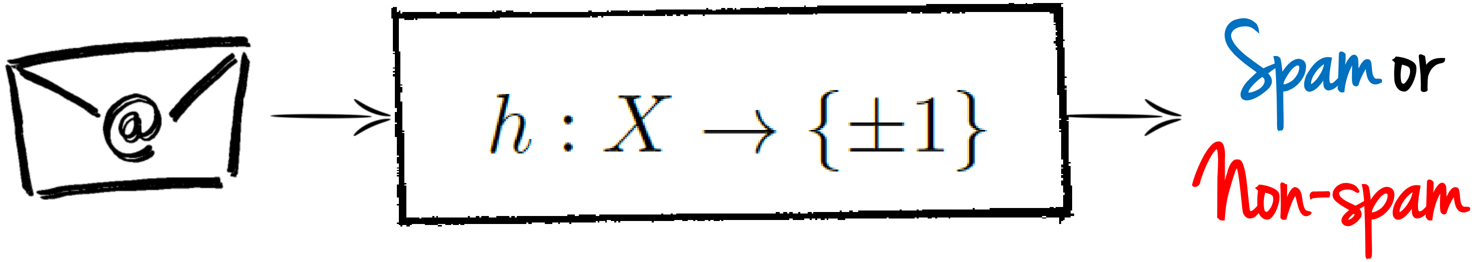


$$h : X \to \{\pm 1\}$$

Spam or Non-spam

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

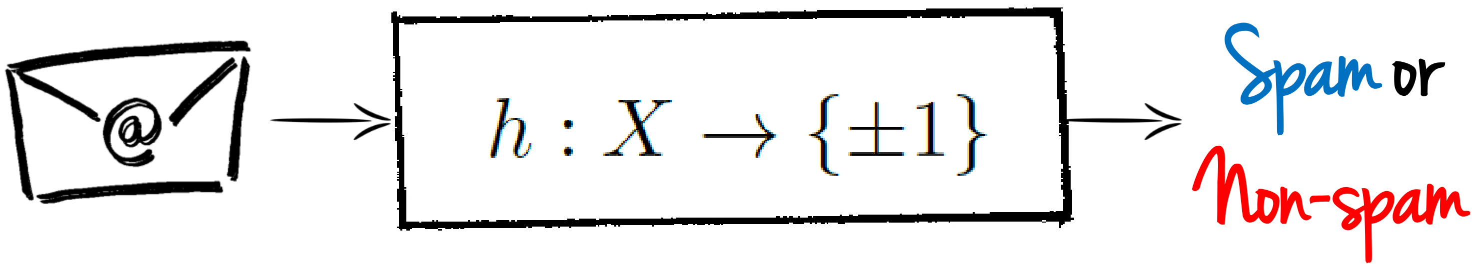
Binary Classification



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

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

Binary Classification



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

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

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

Binary Classification

$\longrightarrow$ $\boxed{h : X \to \{\pm 1\}}$ $\longrightarrow$ Spam or Non-spam

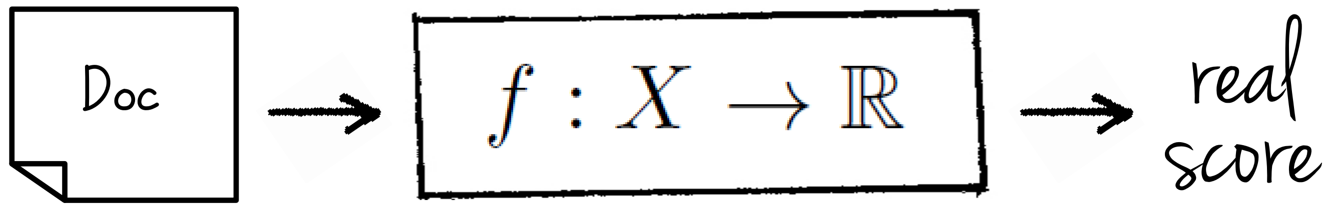
$$S = ((x_1, y_1), \ldots, (x_n, y_n)) \text{ drawn iid from } D$$

0-1 Consistency

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

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

Bipartite Ranking

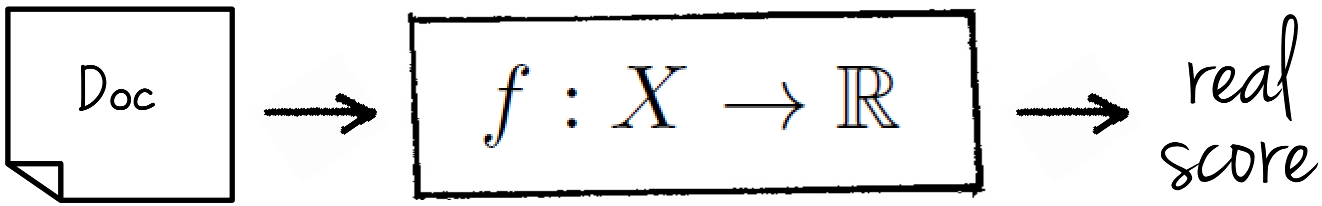


Bipartite Ranking

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

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

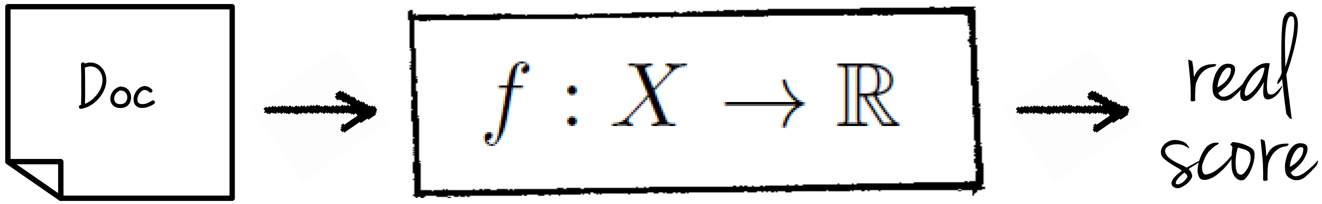
Bipartite Ranking



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

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

Bipartite Ranking

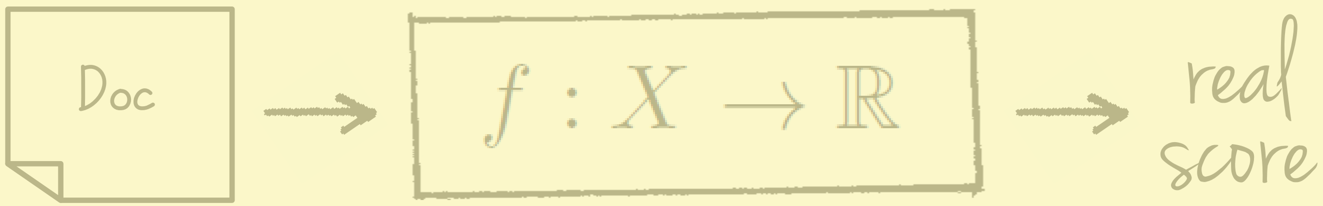


$$\mathrm{er}_D^{\mathrm{rank}}[f] = \mathbf{E}\left[\mathbf{1}\big((y - y')(f(x) - f(x')) < 0\big) \,\middle|\, y \neq y'\right]$$

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

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

Bipartite Ranking



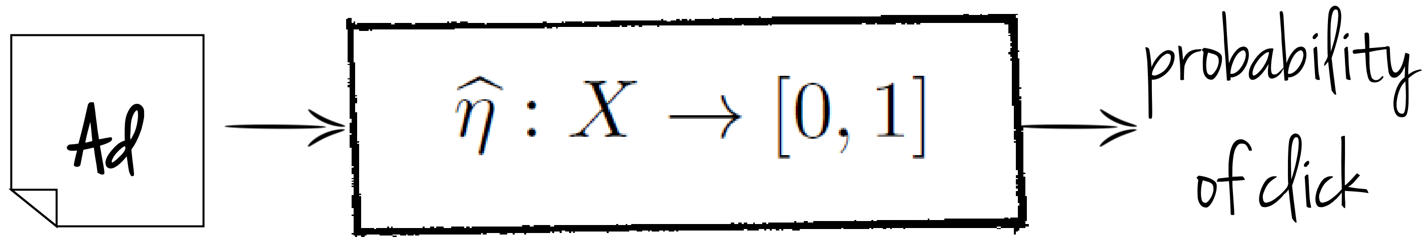
$$S = ((x_1, y_1), \ldots, (x_n, y_n)) \text{ drawn iid from } D$$

Ranking Consistency

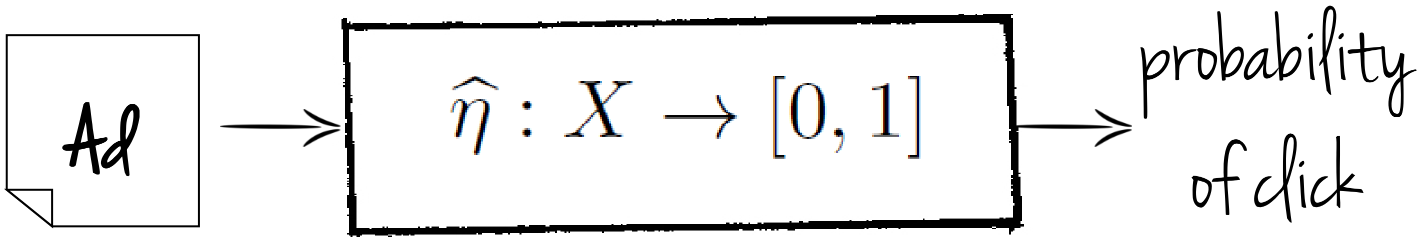
$$\mathrm{regret}_D^{\mathrm{rank}}[f_S] \xrightarrow{P} 0$$

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

Binary Class Probability Estimation

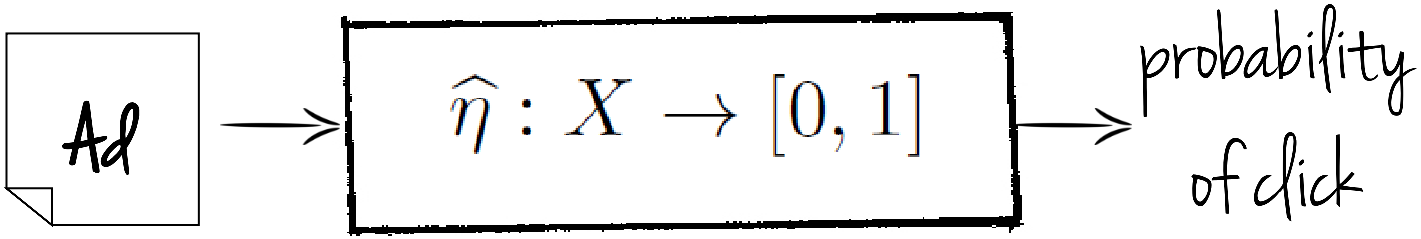


Binary Class Probability Estimation



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

Binary Class Probability Estimation



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

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

Binary Class Probability Estimation

Ad $\longrightarrow$ $\boxed{\widehat{\eta} : X \to [0, 1]}$ $\longrightarrow$ probability of click

$$S = ((x_1, y_1), \ldots, (x_n, y_n)) \text{ drawn iid from } D$$

CPE Consistency

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

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

0-1 Consistency

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

Ranking Consistency

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

CPE Consistency

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

Statistical Consistency

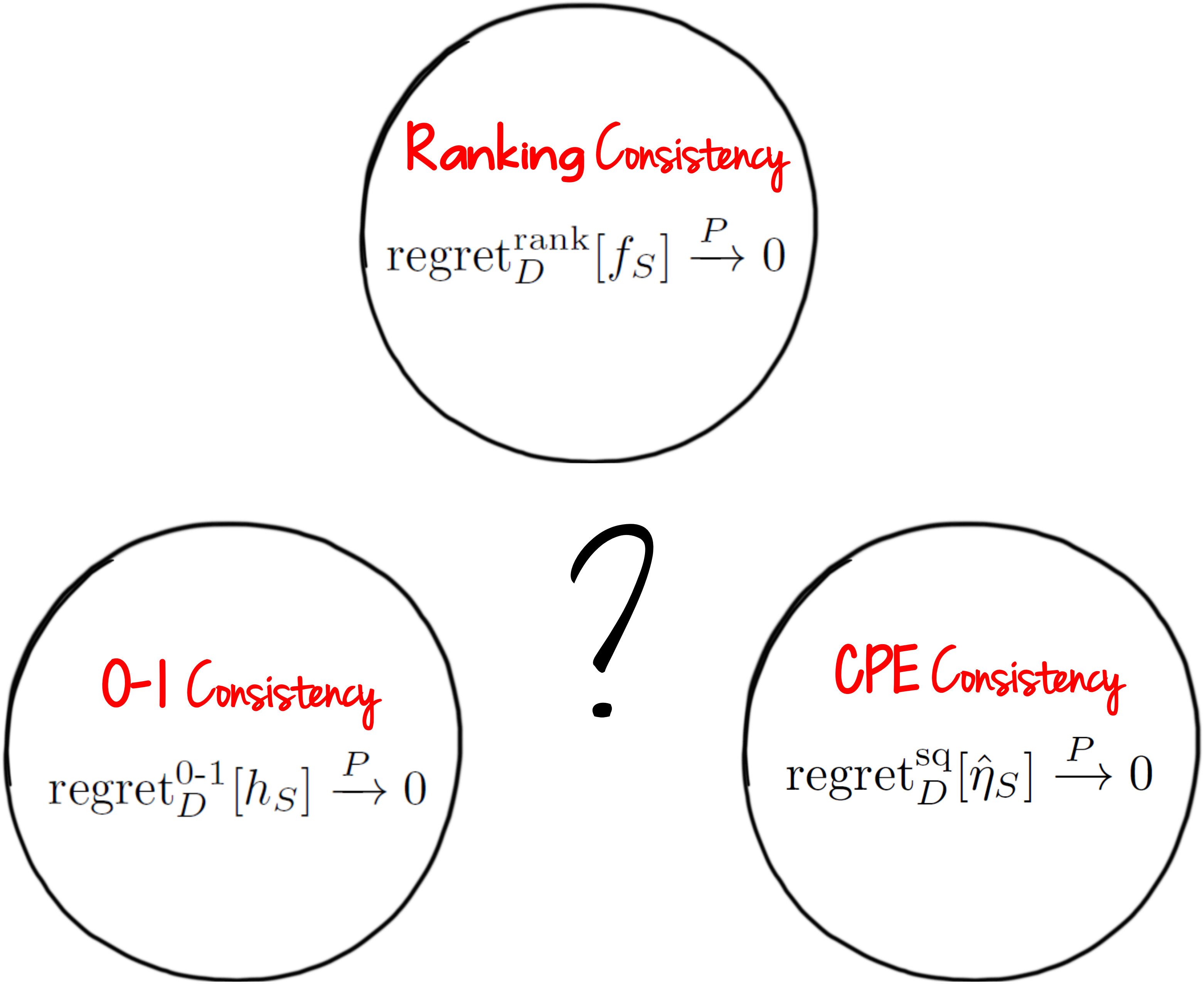
(Surrogate Loss Minimization)

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)

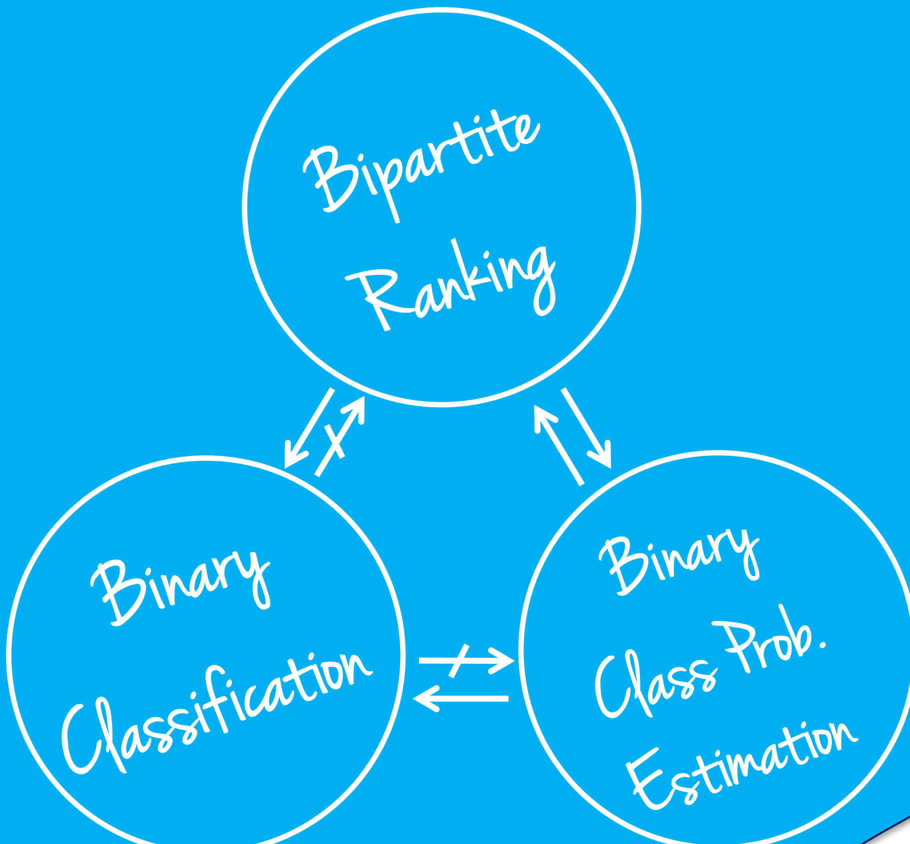

Ranking Consistency
$\mathrm{regret}_D^{\mathrm{rank}}[f_S] \xrightarrow{P} 0$
?
0-1 Consistency
$\mathrm{regret}_D^{0\text{-}1}[h_S] \xrightarrow{P} 0$
CPE Consistency
$\mathrm{regret}_D^{\mathrm{sq}}[\hat{\eta}_S] \xrightarrow{P} 0$

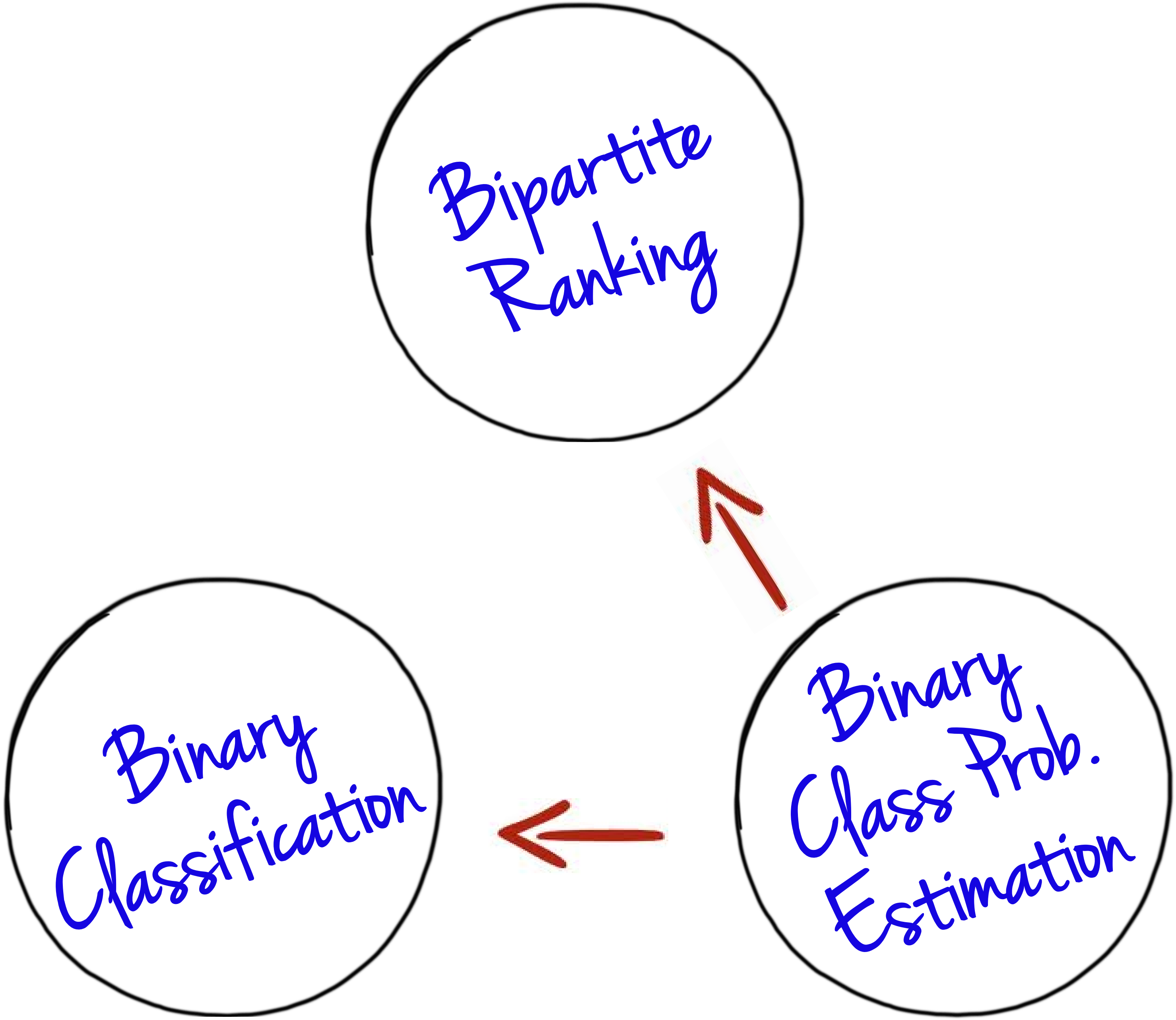
Binary Classification,
Bipartite Ranking,
Binary CPE

Statistical
Consistency

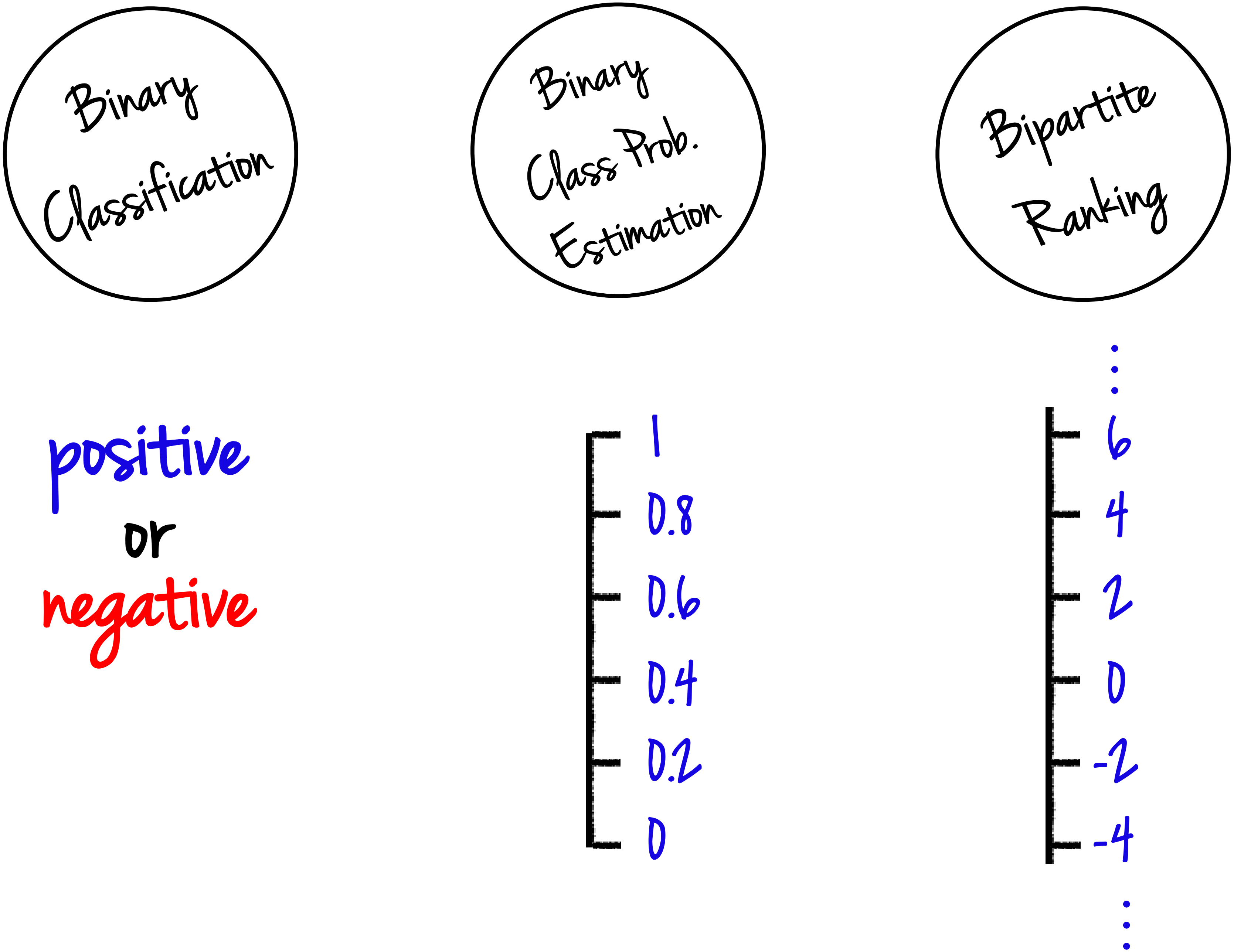
Open
Problems

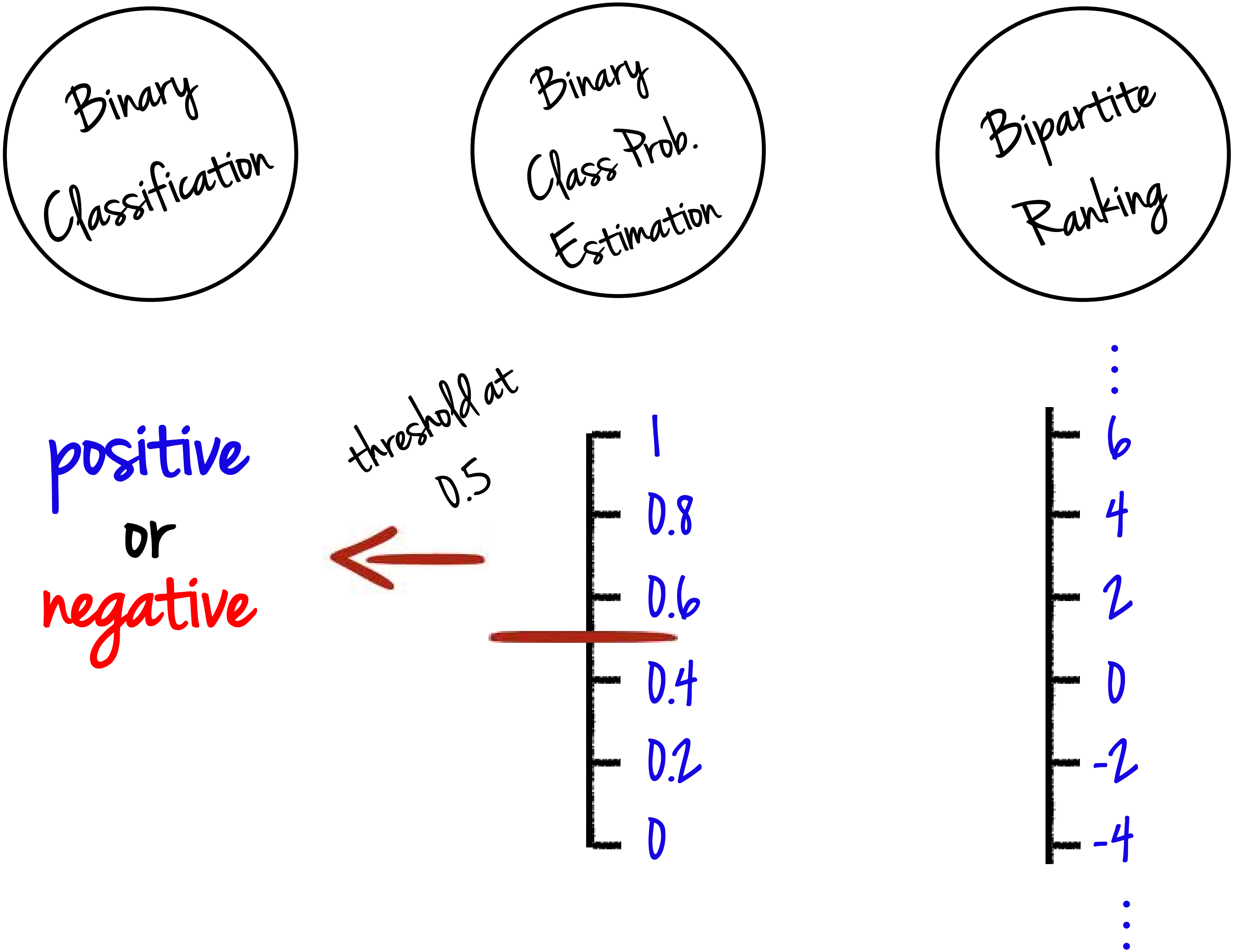
Relationship Between BC, BR, BCPE





Bipartite Ranking
Binary Classification
Binary Class Prob. Estimation


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

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

L. Devroye, L. Gyorfi, 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.

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

L. Devroye, L. Gyorfi, 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.

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

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

L. Devroye, L. Gyorfi, 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.

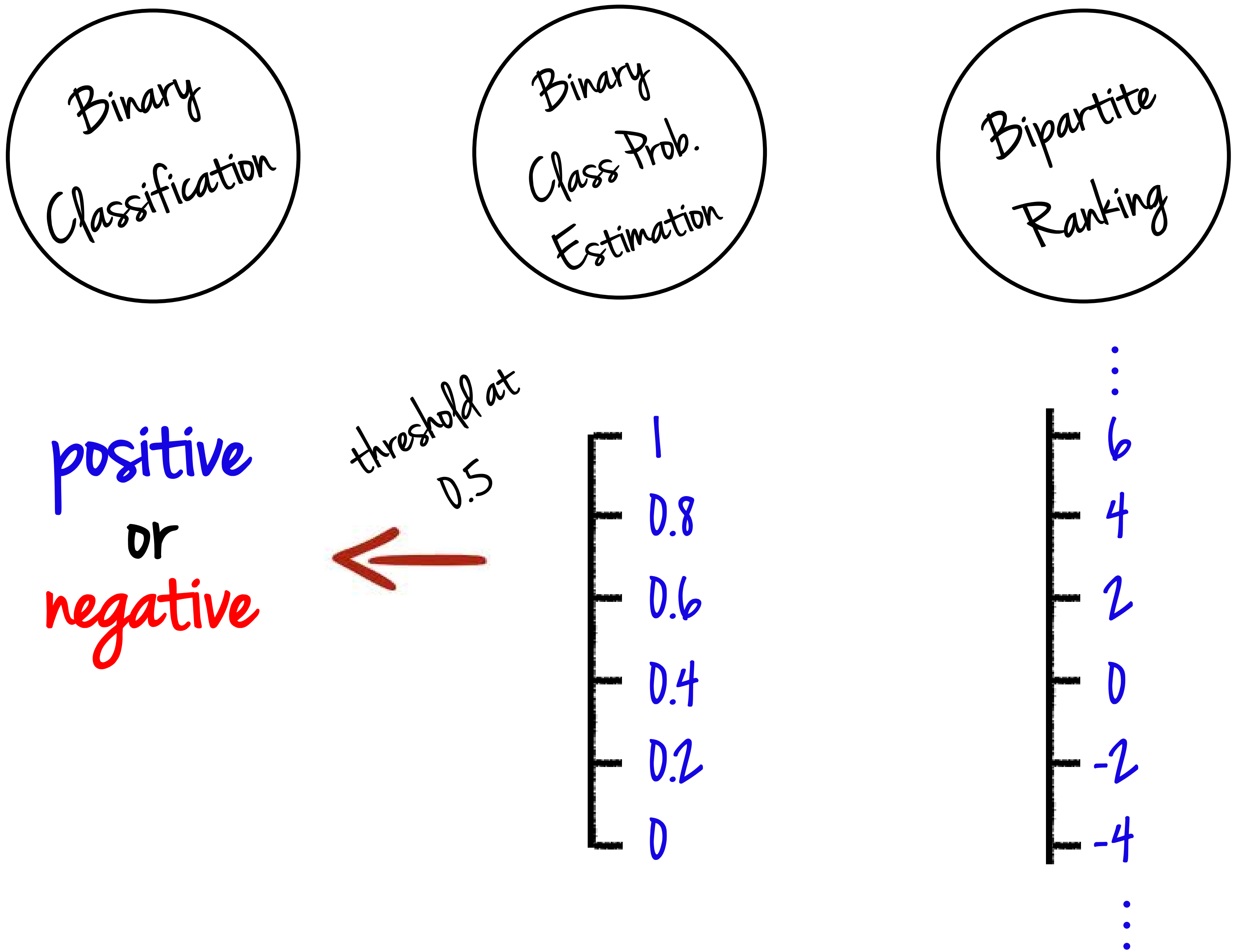
Regret Transfer Bound

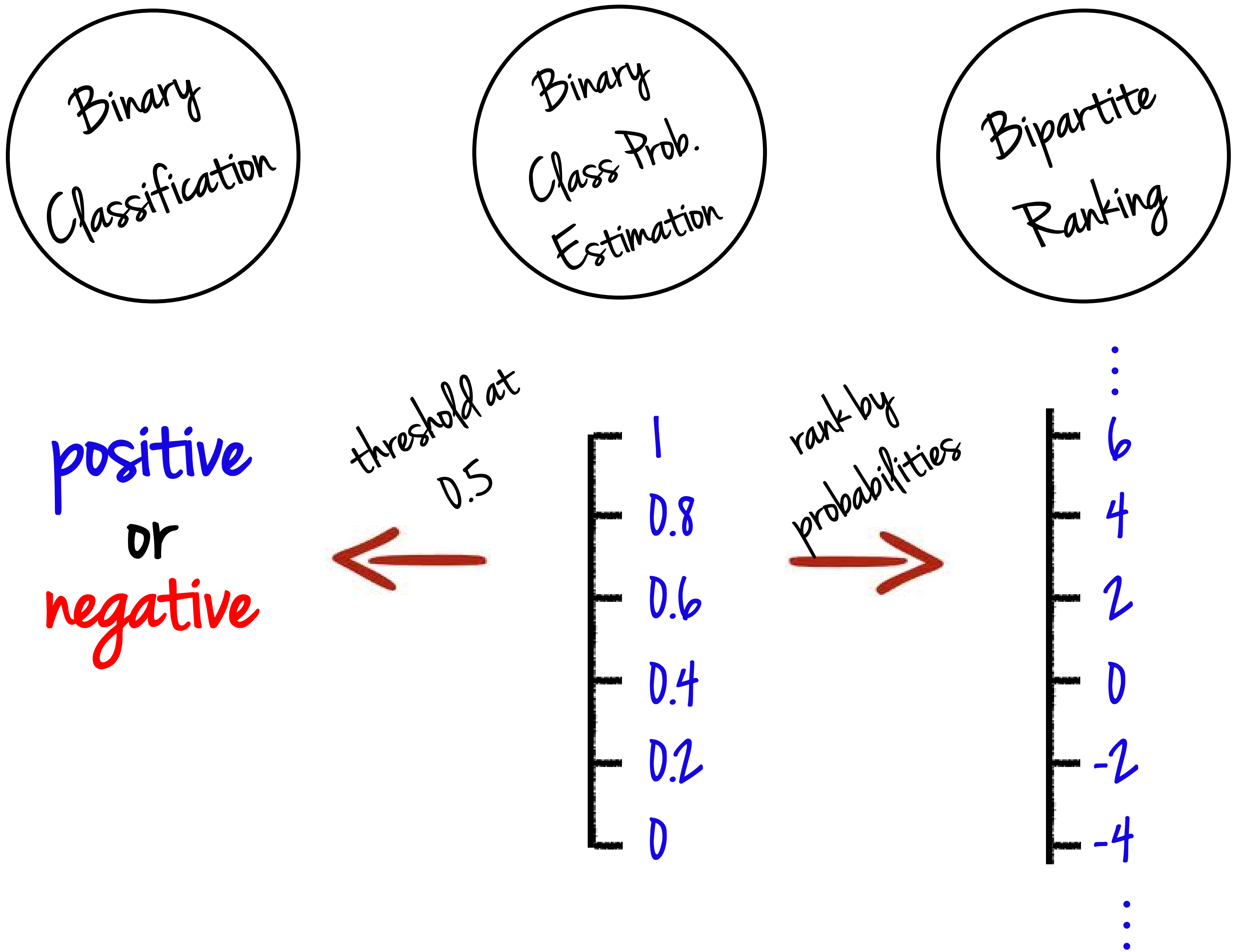
$$\boxed{\mathrm{regret}_D^{0\text{-}1}[\mathrm{sign} \circ (\widehat{\eta} - \tfrac{1}{2})]} \leq \boxed{\sqrt{\mathrm{regret}_D^{\mathrm{sq}}[\,\widehat{\eta}\,]}}$$

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

L. Devroye, L. Gyorfi, 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


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

Regret Transfer Bound

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

S. Agarwal, "Surrogate regret bounds for bipartite ranking via strongly proper losses," arXiv:1207.0268, 2012.

Regret Transfer Bound

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

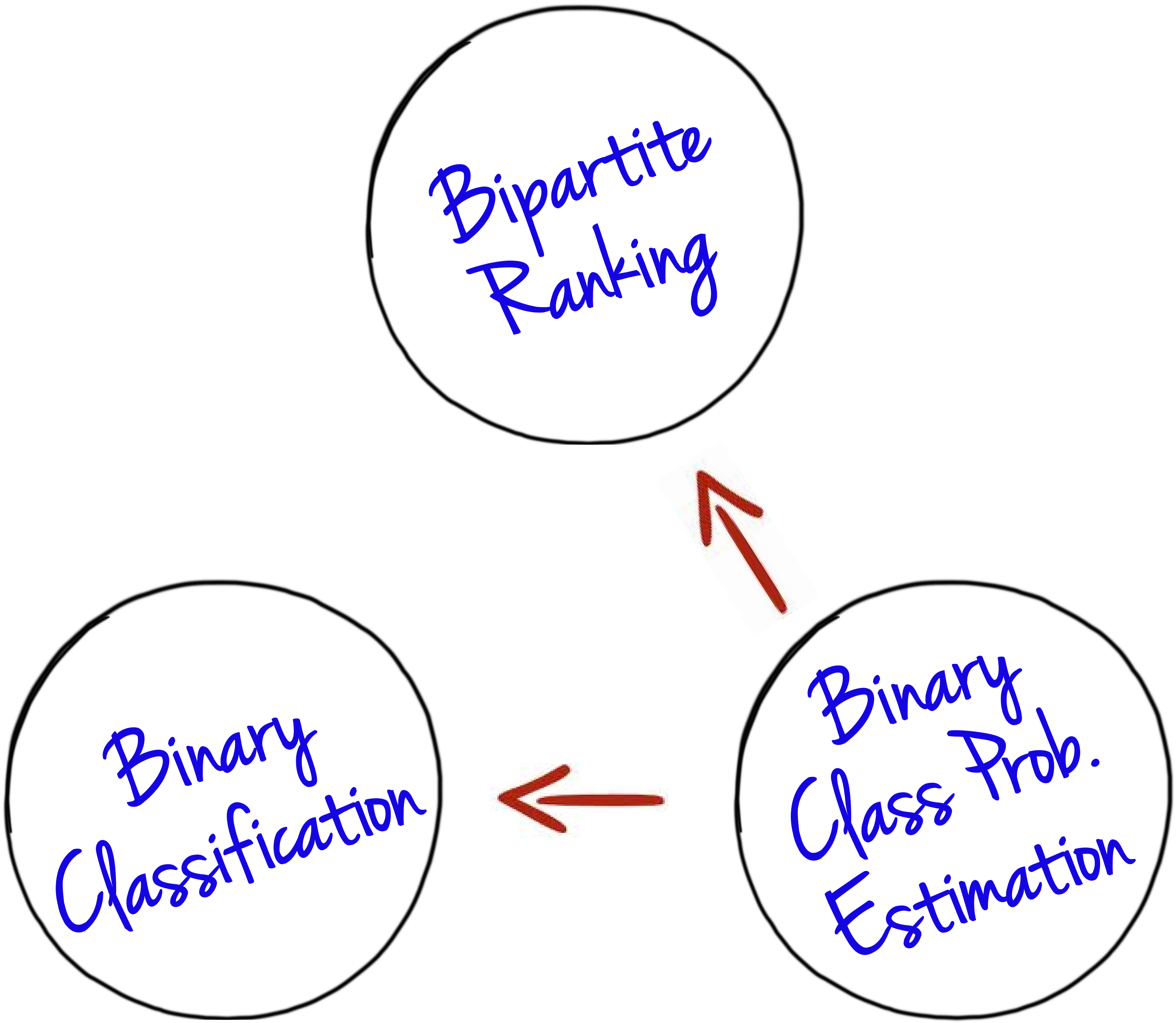
S. Agarwal, “Surrogate regret bounds for bipartite ranking via strongly proper losses,” arXiv:1207.0268, 2012.

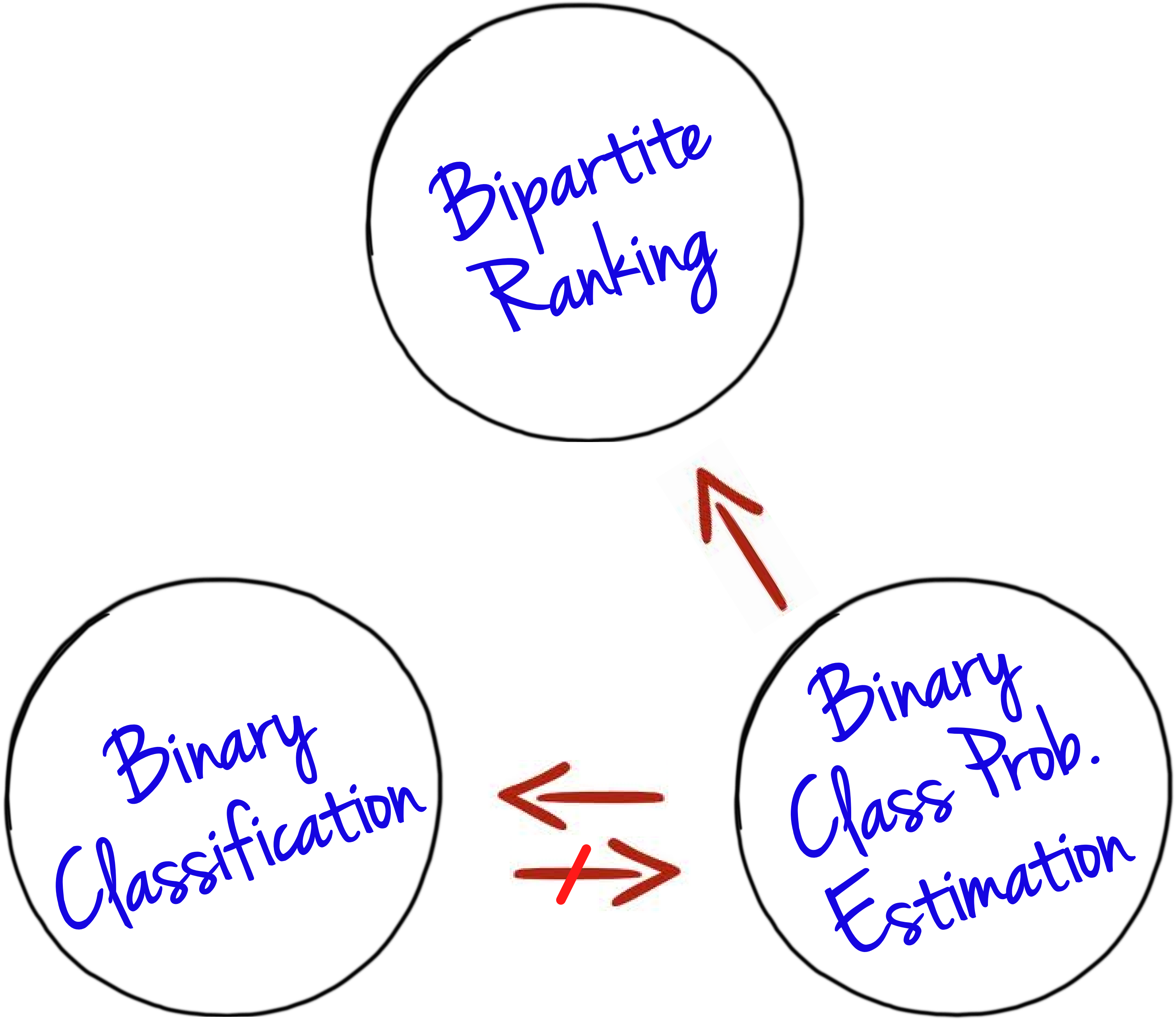
Regret Transfer Bound

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

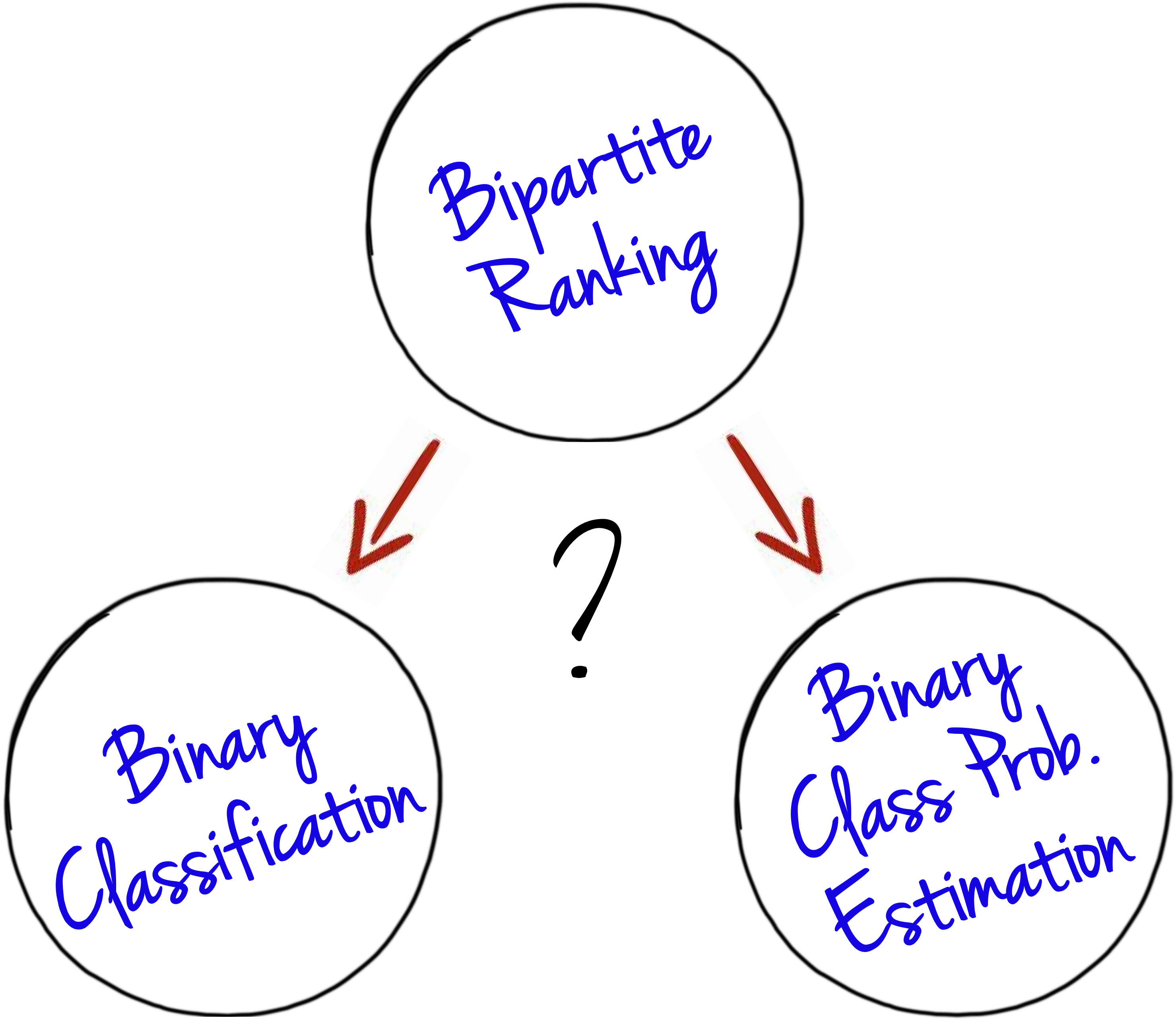
$$\xrightarrow{P} 0 \qquad\qquad \xrightarrow{P} 0$$

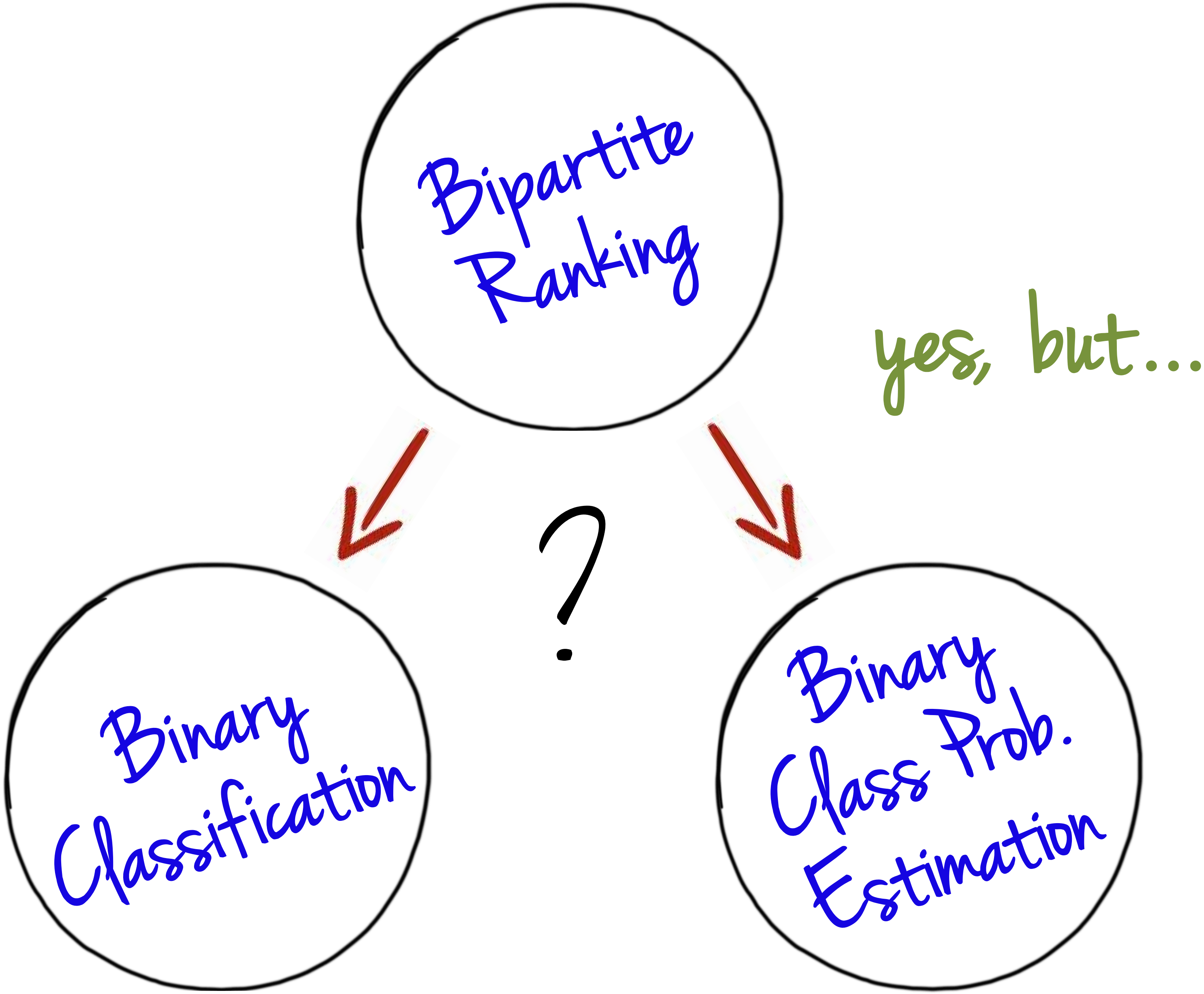
S. Agarwal, "Surrogate regret bounds for bipartite ranking via strongly proper losses," arXiv:1207.0268, 2012.

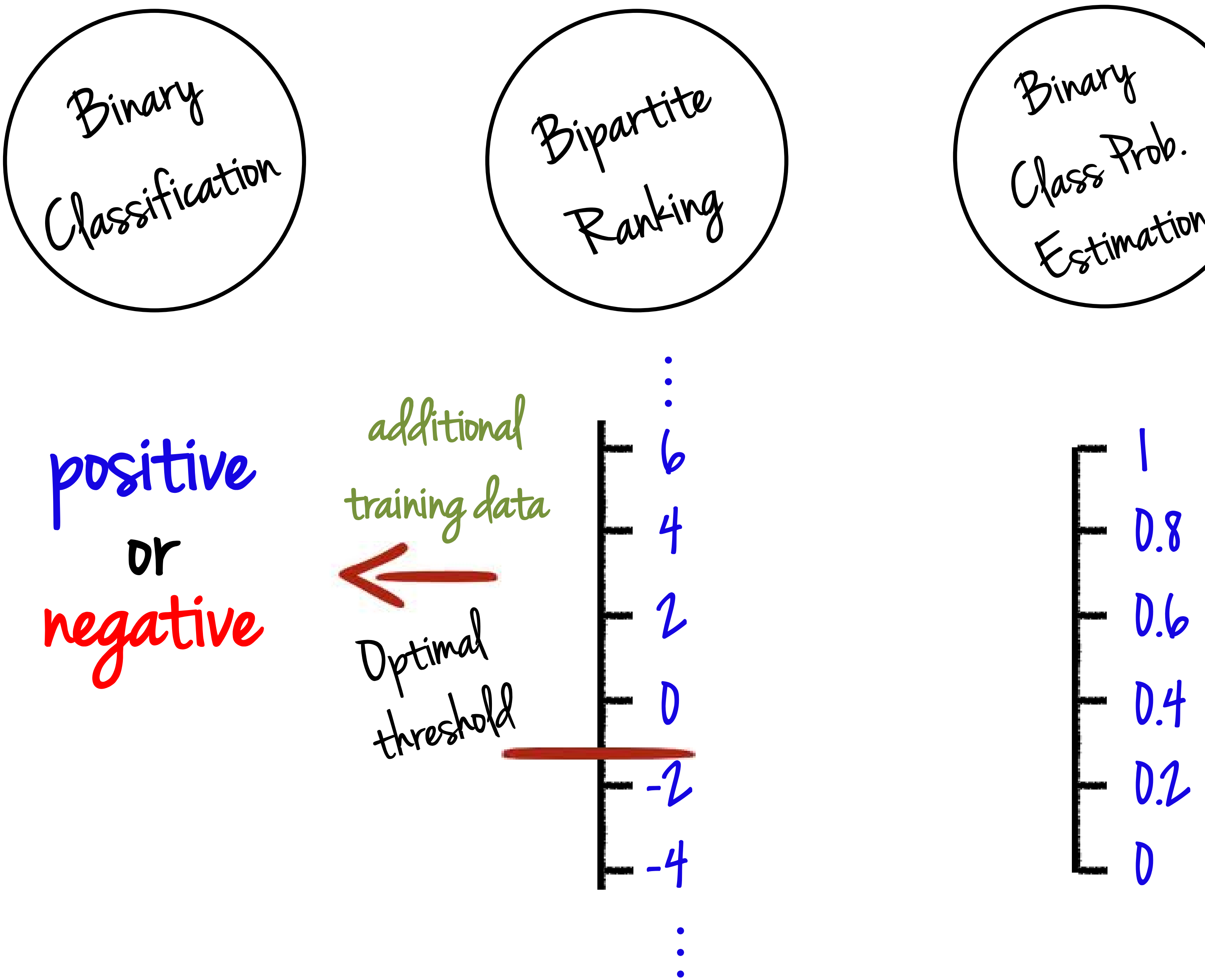

Bipartite Ranking
Binary Classification
Binary Class Prob. Estimation



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.


Bipartite Ranking
?
Binary Classification
Binary Class Prob. Estimation


Bipartite Ranking
yes, but...
?
Binary Classification
Binary Class Prob. Estimation


Binary Classification
Bipartite Ranking
Binary Class Prob. Estimation
positive
or
negative
additional training data
Optimal threshold
6
4
2
0
-2
-4
1
0.8
0.6
0.4
0.2
0

Bipartite Ranking $\longrightarrow$ Binary Classification

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

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

Bipartite Ranking $\longrightarrow$ Binary Classification

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

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

Regret Transfer Bound

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

Bipartite Ranking → Binary Classification

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

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

Regret Transfer Bound

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

$$\to 0$$

Bipartite Ranking $\longrightarrow$ Binary Classification

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

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

Regret Transfer Bound

$$\underbrace{\boxed{\mathrm{regret}_D^{0\text{-}1}\left[\mathrm{sign}\circ(f-t^*)\right]}}_{\to 0} \;\le\; C'\sqrt{\underbrace{\boxed{\mathrm{regret}_D^{\mathrm{rank}}[f]}}_{\to 0}}$$

Bipartite Ranking → Binary Classification

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

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

(Weak) Regret Transfer Bound

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

dependence on distribution

Bipartite Ranking → Binary Classification

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

$$S = ((x_1, y_1), \ldots, (x_n, y_n)) \text{ drawn iid from } D$$

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

Sample-Based (Weak) Regret Transfer Bound

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

Bipartite Ranking $\longrightarrow$ Binary Classification

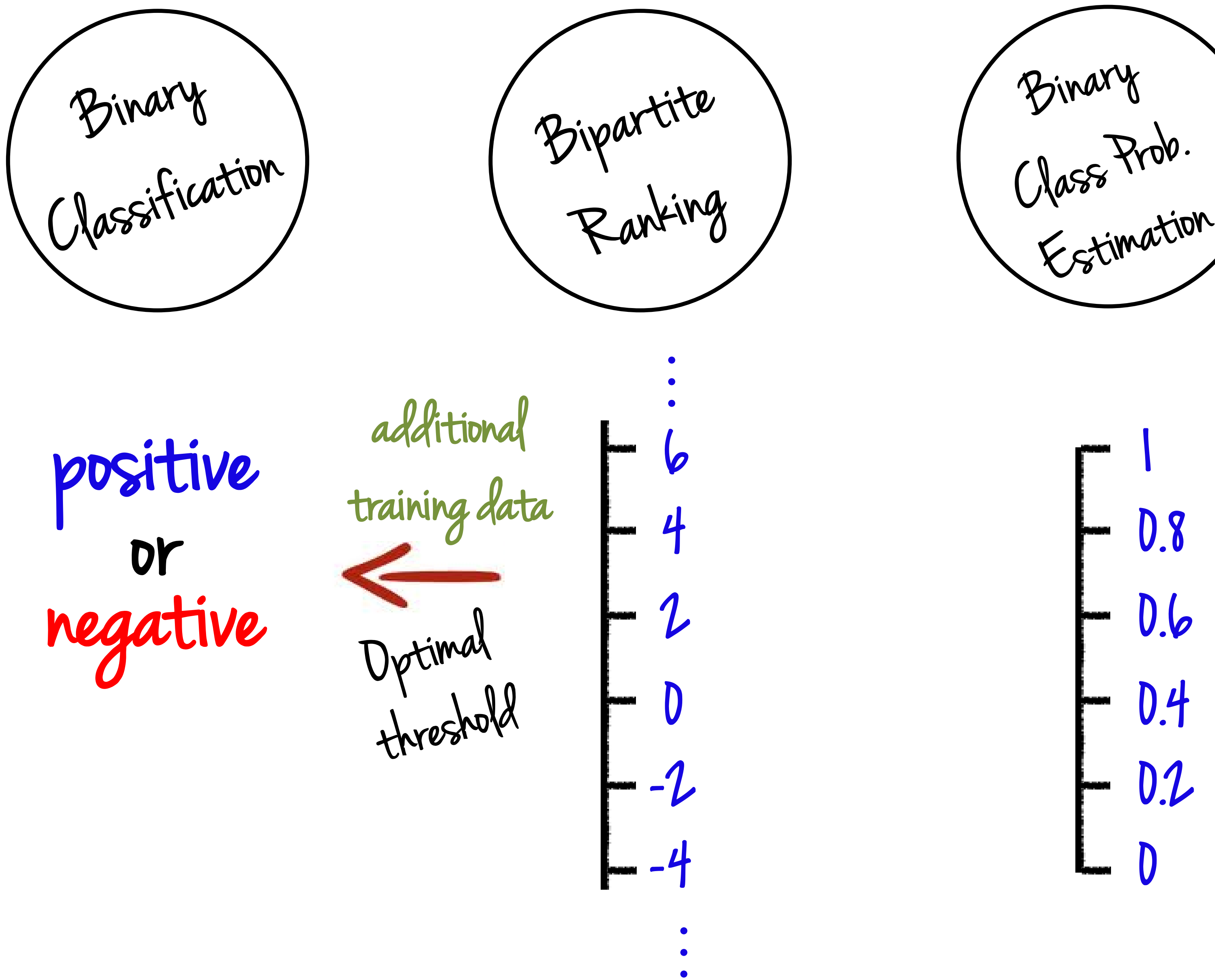
$$f : X \to \mathbb{R}$$

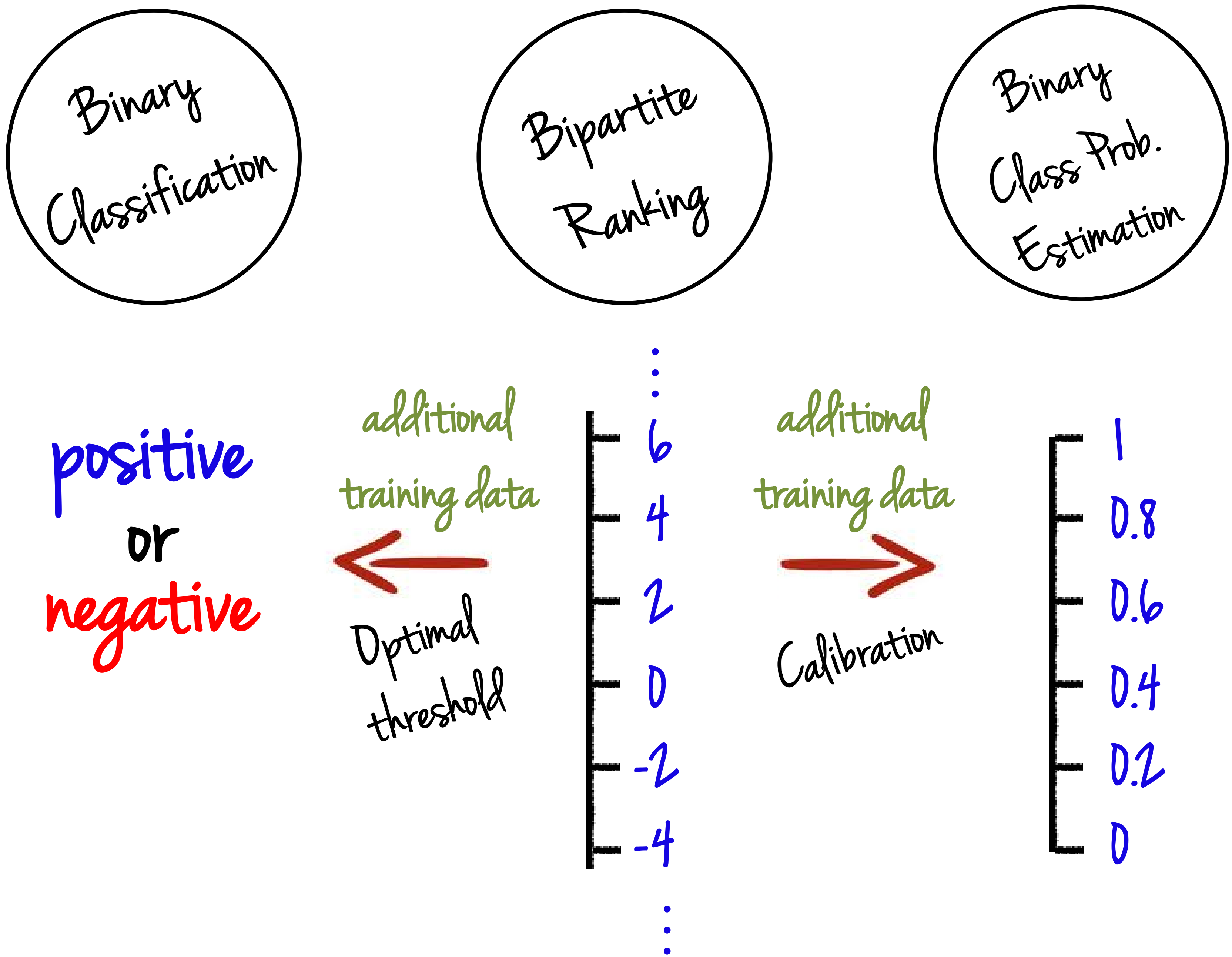
$$S = ((x_1, y_1), \ldots, (x_n, y_n)) \text{ drawn iid from } D$$

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

Sample-Based (Weak) Regret Transfer Bound

$$\mathrm{regret}_D^{0\text{-}1}\left[\mathrm{sign} \circ (f - \hat{t})\right] \;\leq\; C' \sqrt{\mathrm{regret}_D^{\mathrm{rank}}[f]} \;+ \underbrace{\boxed{O\left(\sqrt{\frac{\ln(n)}{n}}\right)}}_{\to\, 0}$$


Binary Classification
Bipartite Ranking
Binary Class Prob. Estimation
positive
or
negative
additional training data
Optimal threshold
6
4
2
0
-2
-4
1
0.8
0.6
0.4
0.2
0



Binary Classification
Bipartite Ranking
Binary Class Prob. Estimation
positive
or
negative
additional training data
Optimal threshold
additional training data
Calibration
6
4
2
0
-2
-4
1
0.8
0.6
0.4
0.2
0

Bipartite Ranking $\longrightarrow$ Binary CPE

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

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

monotonically increasing functions

Bipartite Ranking $\longrightarrow$ Binary CPE

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

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

monotonically increasing functions

(Weak) Regret Transfer Bound

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

Bipartite Ranking $\longrightarrow$ Binary CPE

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

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

monotonically increasing functions

(Weak) Regret Transfer Bound

$$\underbrace{\boxed{\mathrm{regret}_D^{\mathrm{sq}}[\mathrm{Cal} \circ f]}}_{\to\, 0} \;\leq\; C'' \sqrt{\underbrace{\boxed{\mathrm{regret}_D^{\mathrm{rank}}[f]}}_{\to\, 0}}$$

Bipartite Ranking → Binary CPE

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

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

monotonically increasing functions

(Weak) Regret Transfer Bound

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

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

dependence on distribution

Bipartite Ranking $\rightarrow$ Binary CPE

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

$$S = ((x_1, y_1), \ldots, (x_n, y_n)) \text{ drawn iid from } D$$

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

Sample-Based (Weak) Regret Transfer Bound

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

Bipartite Ranking → Binary CPE

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

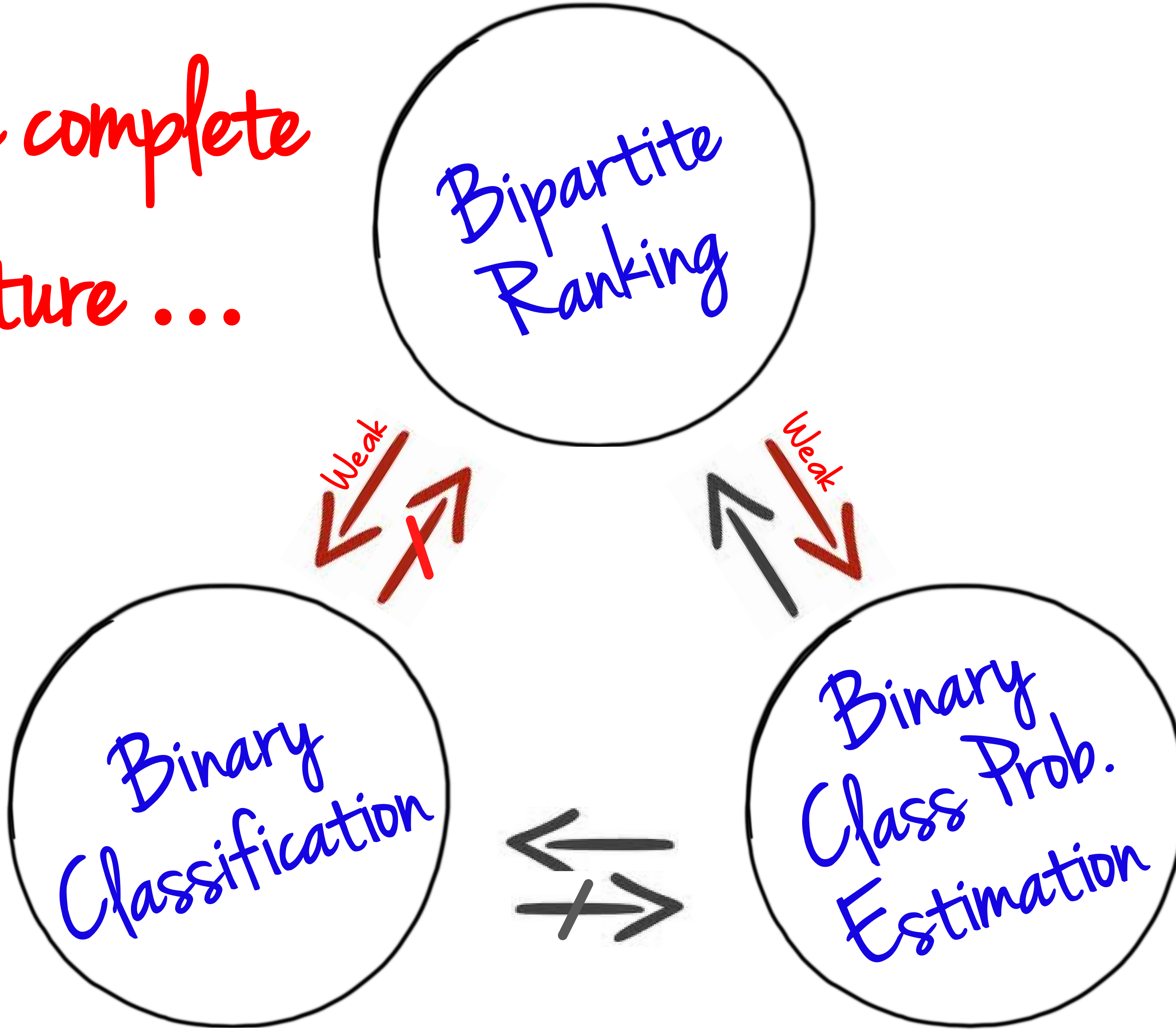
$$S = ((x_1, y_1), \ldots, (x_n, y_n)) \text{ drawn iid from } D$$

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

Sample-Based (Weak) Regret Transfer Bound

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

the complete picture ...



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

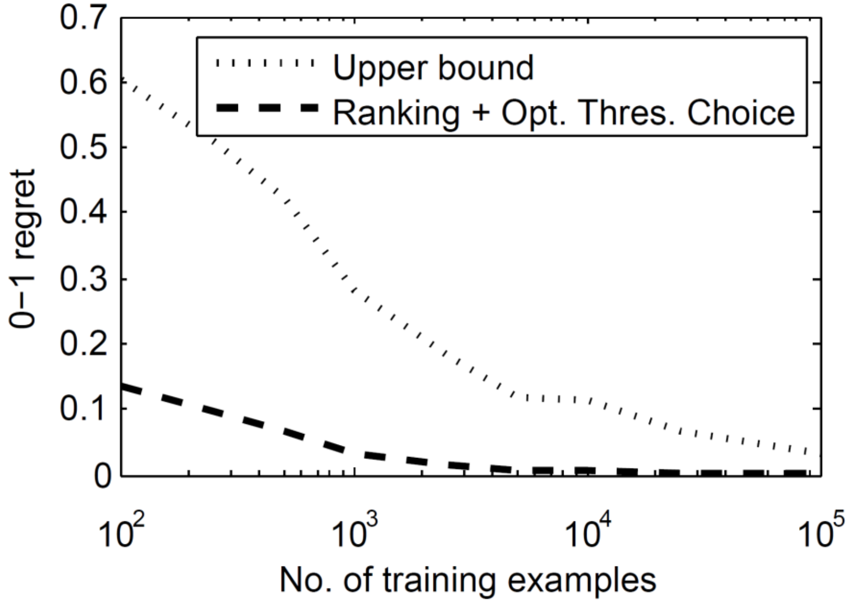
Experiments: Synthetic Data

Validating Regret Transfer Bounds

Experiments: Synthetic Data

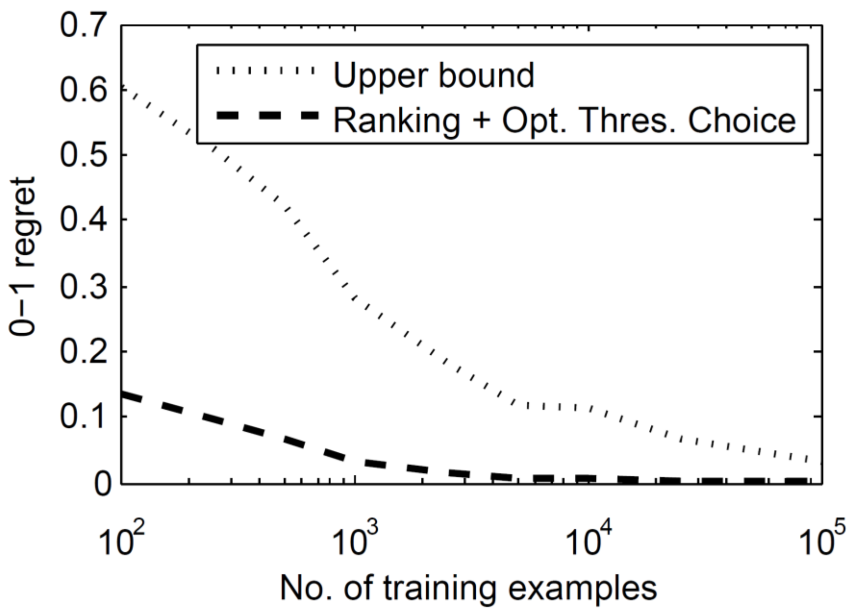
Validating Regret Transfer Bounds

Bipartite Ranking → Binary Classification

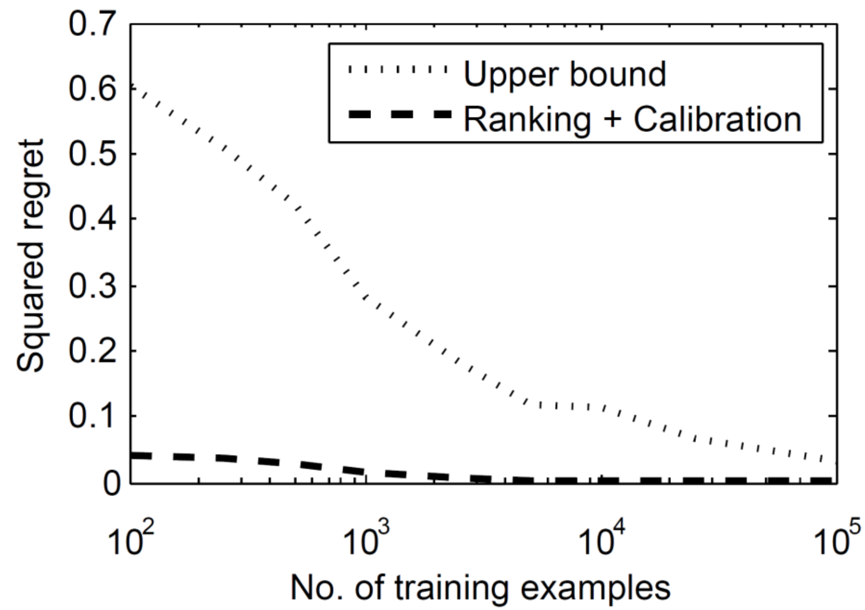


Experiments: Synthetic Data
Validating Regret Transfer Bounds

Bipartite Ranking → Binary Classification



Bipartite Ranking → Binary CPE



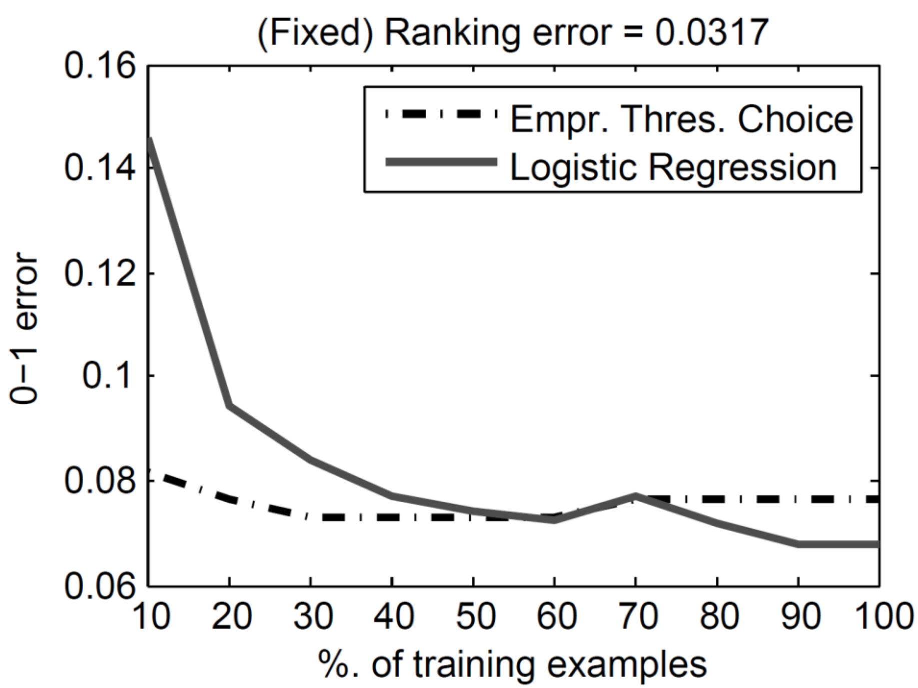
Experiments: Real Data

Comparison with Learning from Scratch

Experiments: Real Data

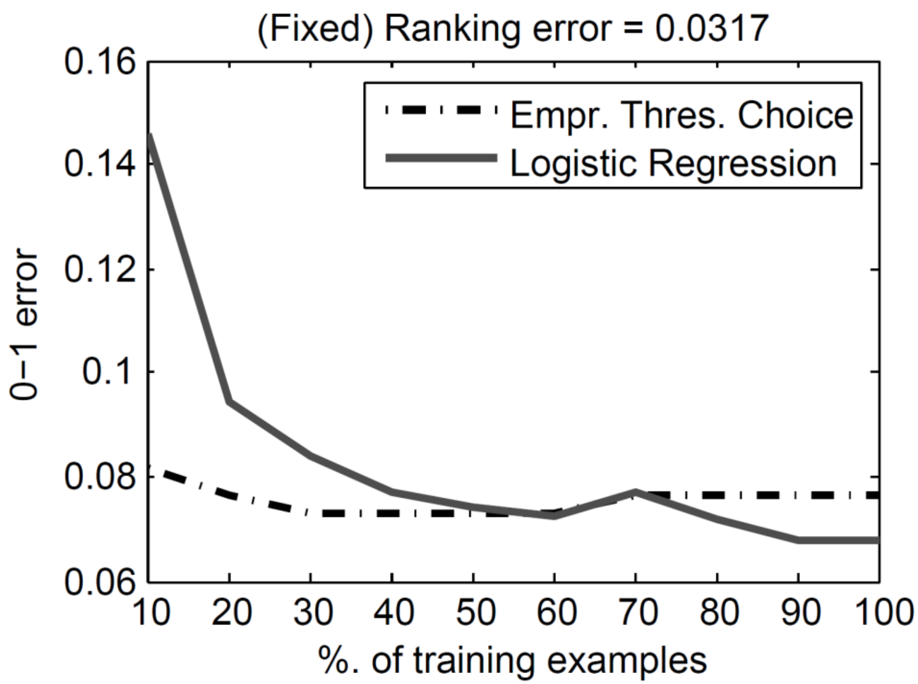
Comparison with Learning from Scratch

Bipartite Ranking → Binary Classification

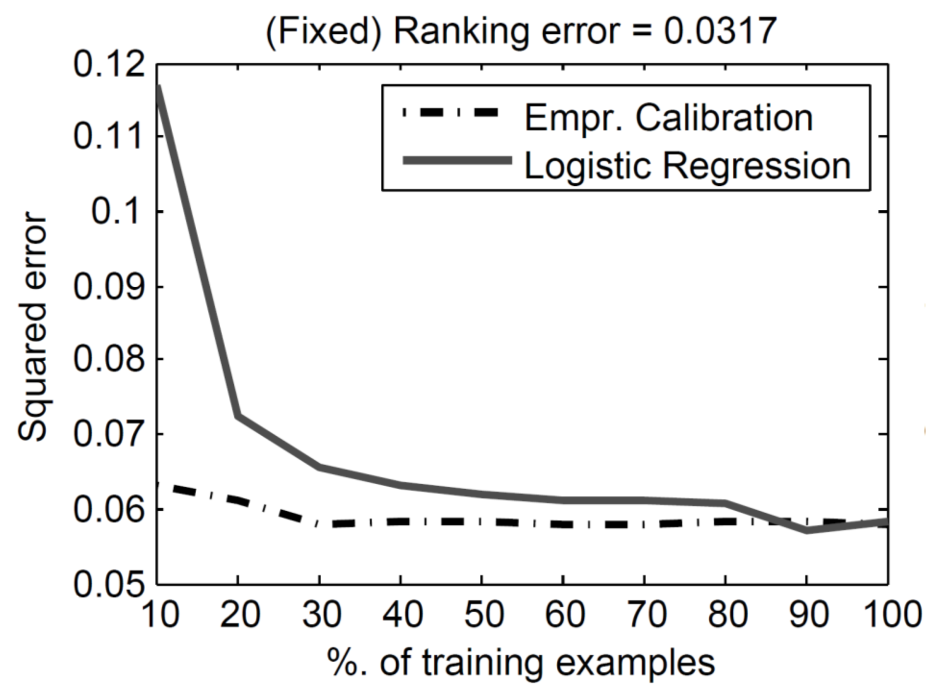


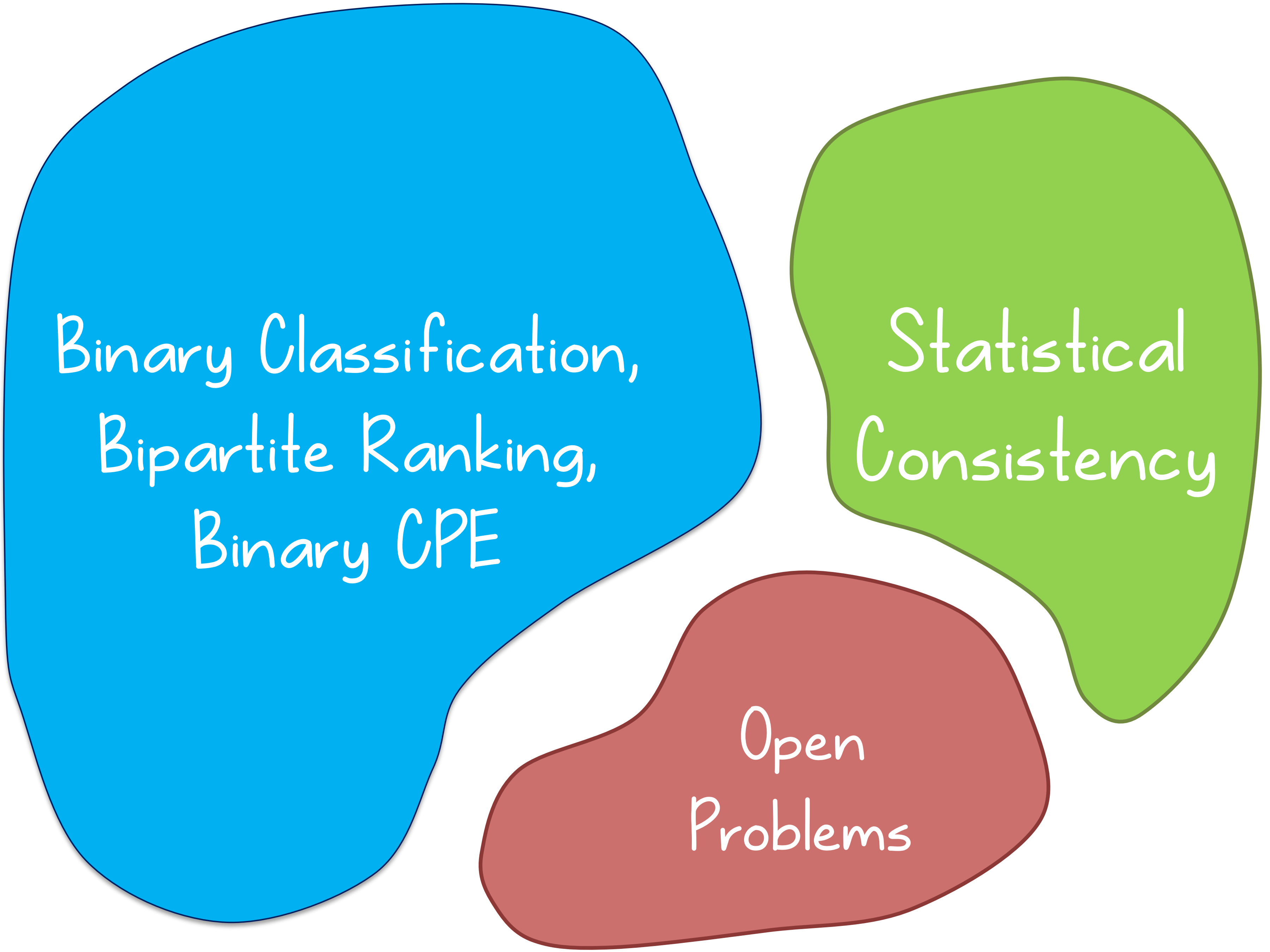
Experiments: Real Data
Comparison with Learning from Scratch

Bipartite Ranking ⟶ Binary Classification

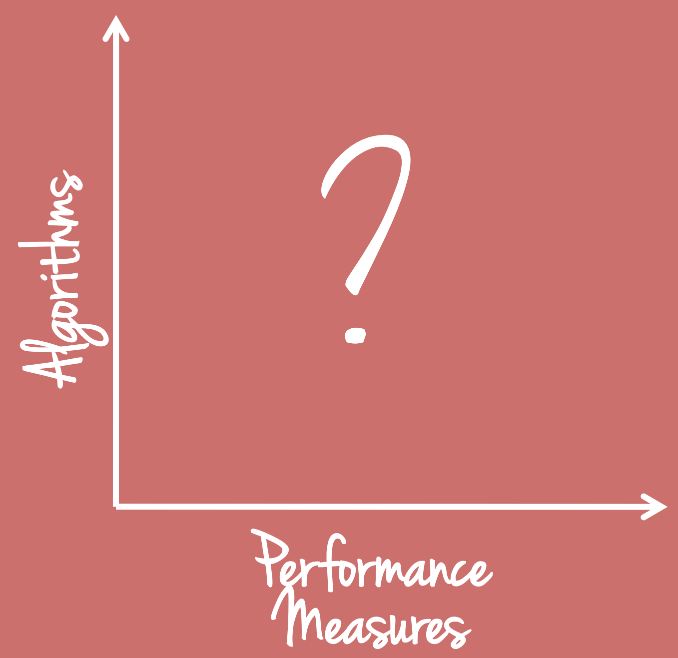


Bipartite Ranking ⟶ Binary CPE




Binary Classification,
Bipartite Ranking,
Binary CPE
Statistical
Consistency
Open
Problems

Open Problems



performance measure?

Binary Classification

0-1 Classification Error

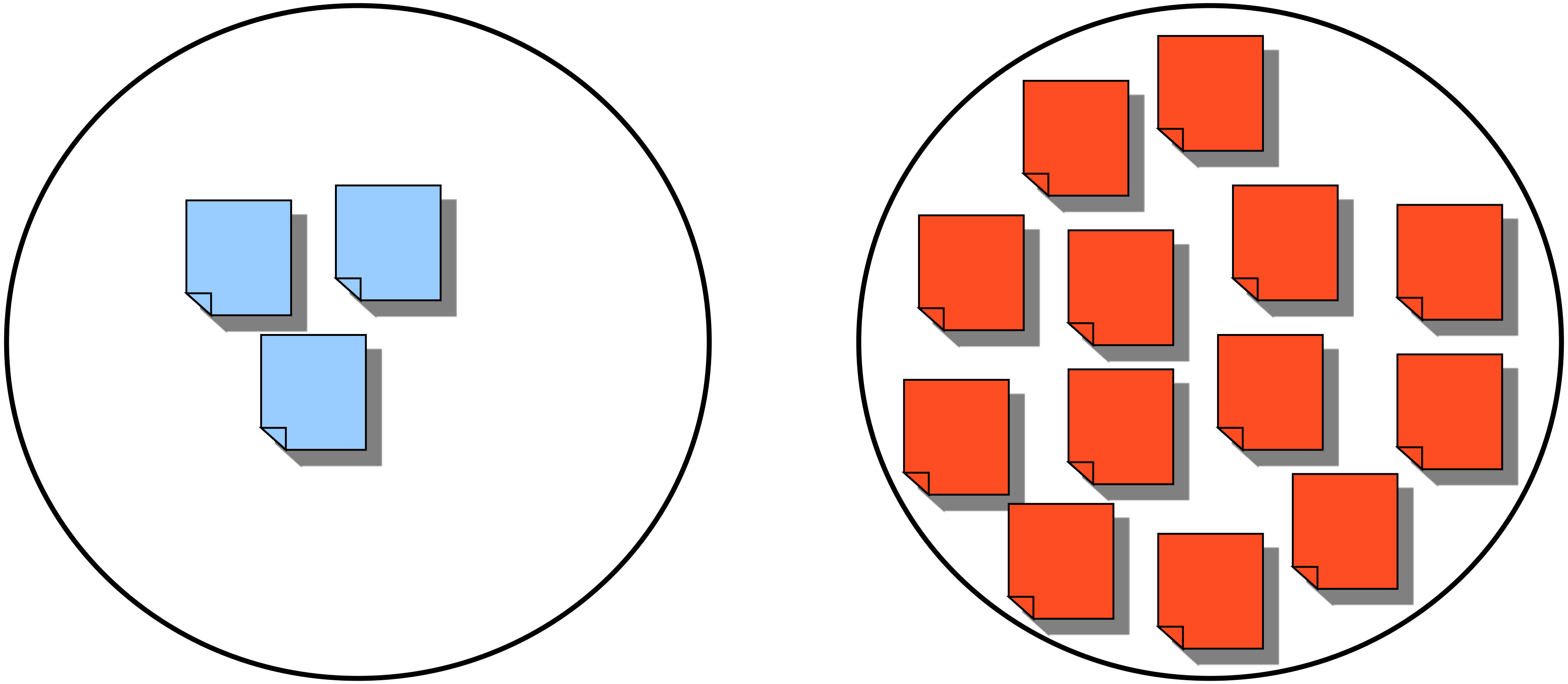
Bipartite Ranking

Pair-wise Accuracy

Binary Class Prob. Estimation

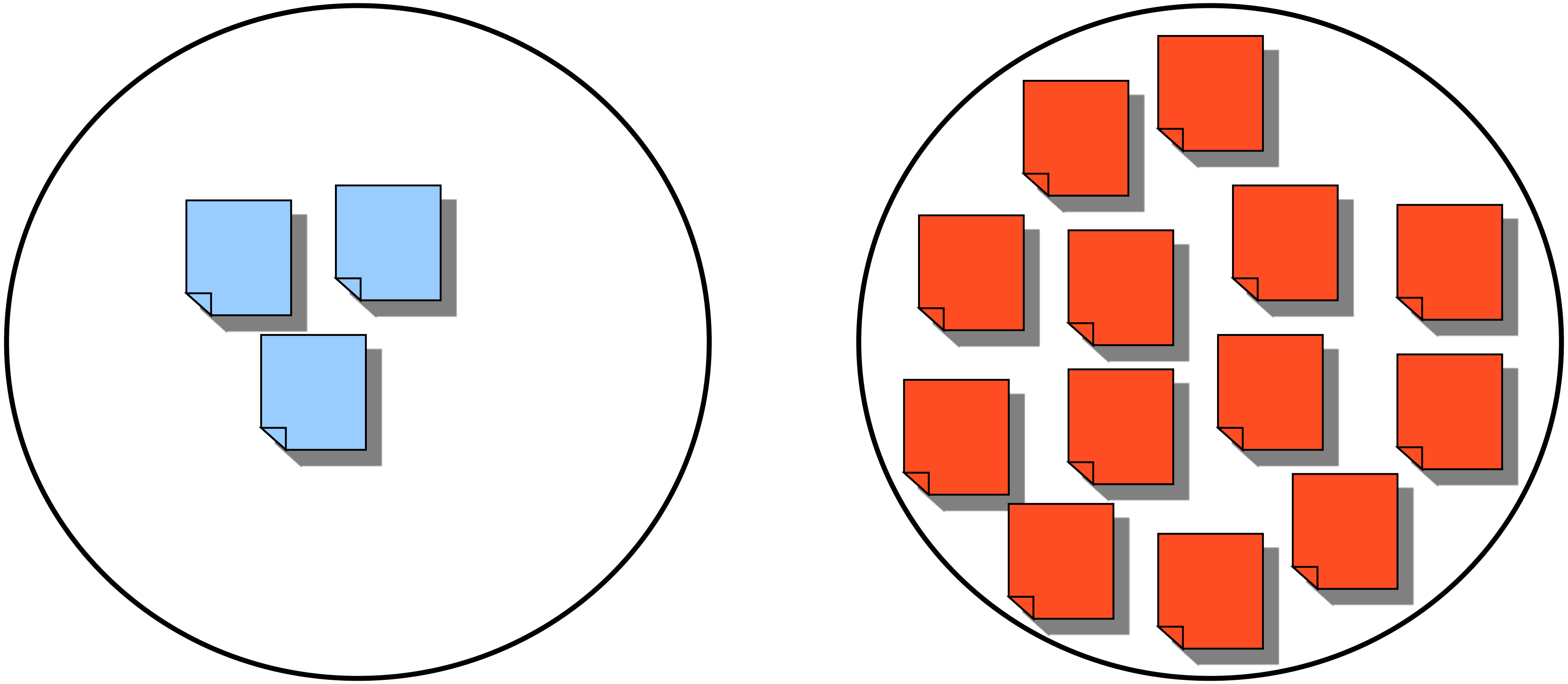
Squared Error

Text Retrieval



Text Retrieval

F-measure



Information Retrieval

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 ...

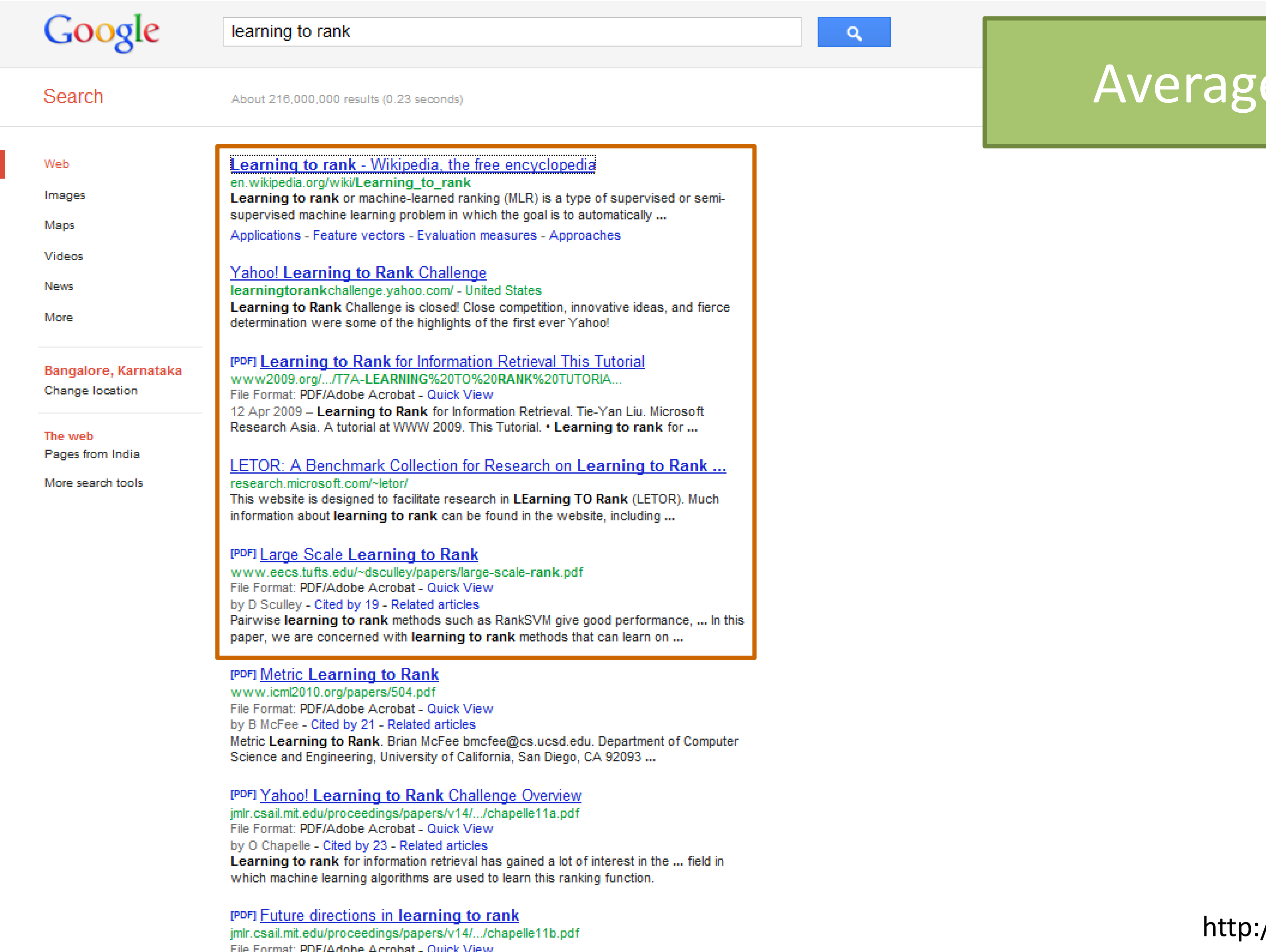
[PDF] Metric **Learning to Rank**
www.icml2010.org/papers/504.pdf
File Format: PDF/Adobe Acrobat - Quick View
by B McFee - Cited by 21 - Related articles
Metric **Learning to Rank**. Brian McFee bmcfee@cs.ucsd.edu. Department of Computer Science and Engineering, University of California, San Diego, CA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

http://www.google.com/

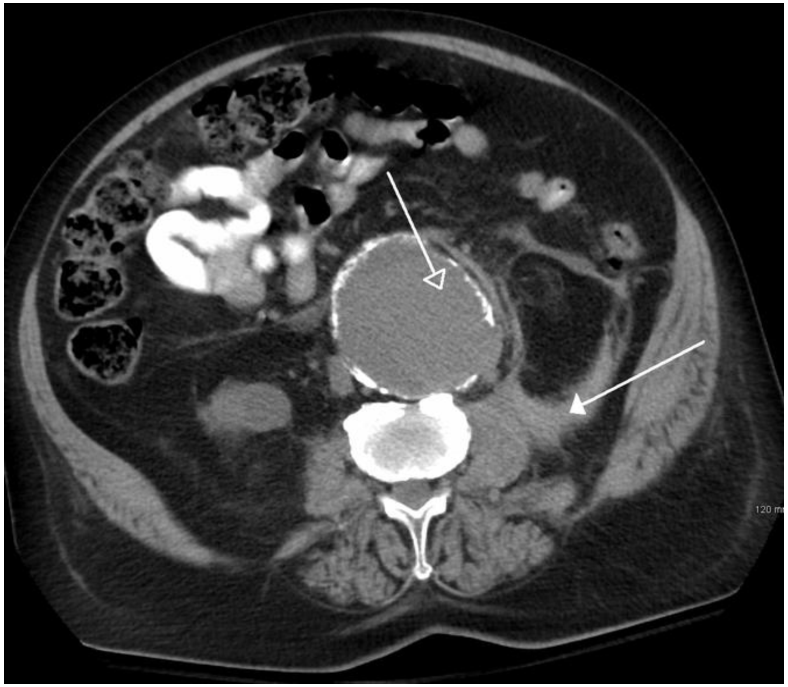
Information Retrieval

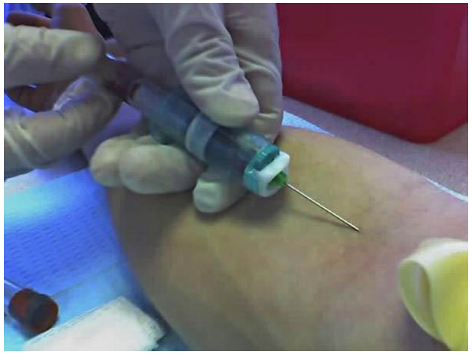


Average Precision

http://www.google.com/

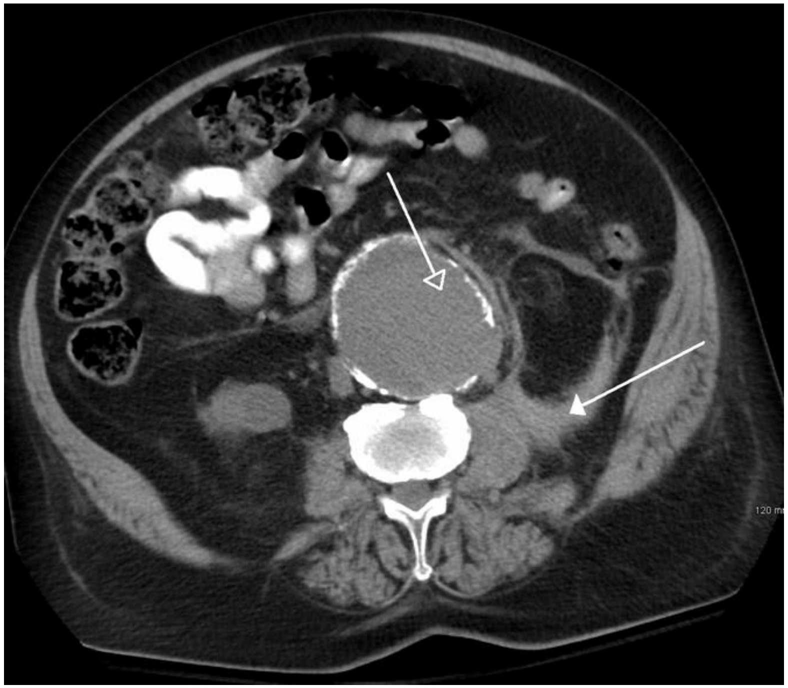
Medical Diagnosis

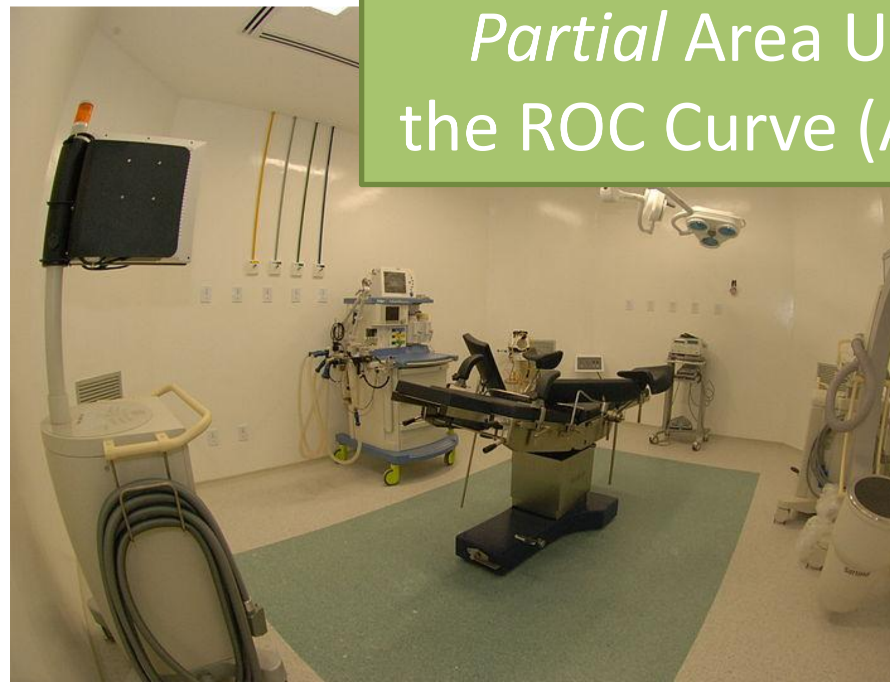




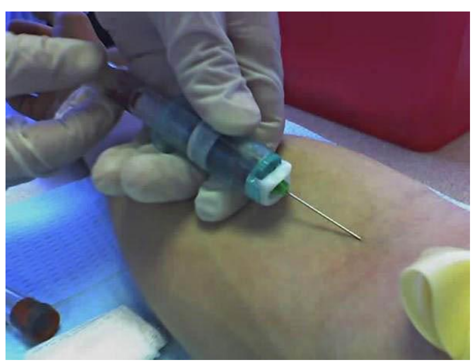
http://en.wikipedia.org/

Medical Diagnosis

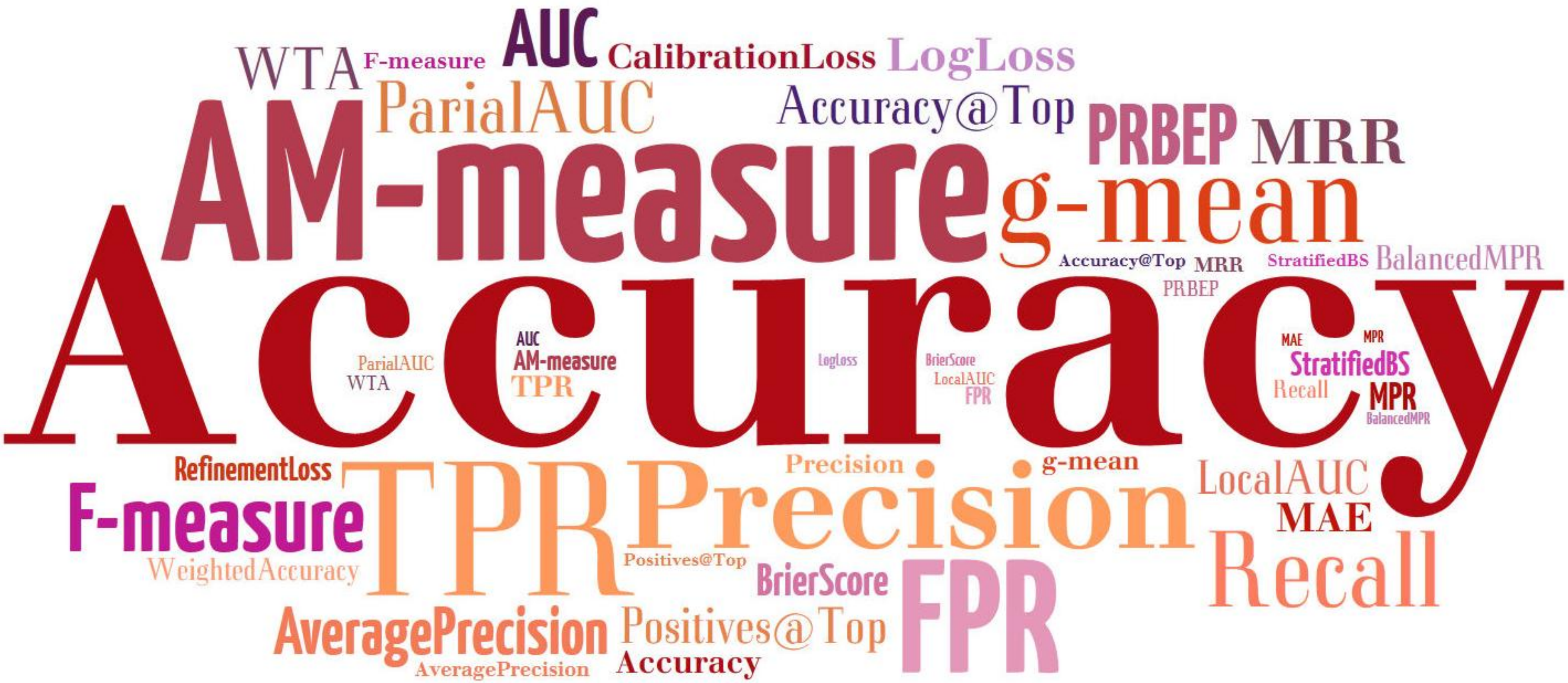




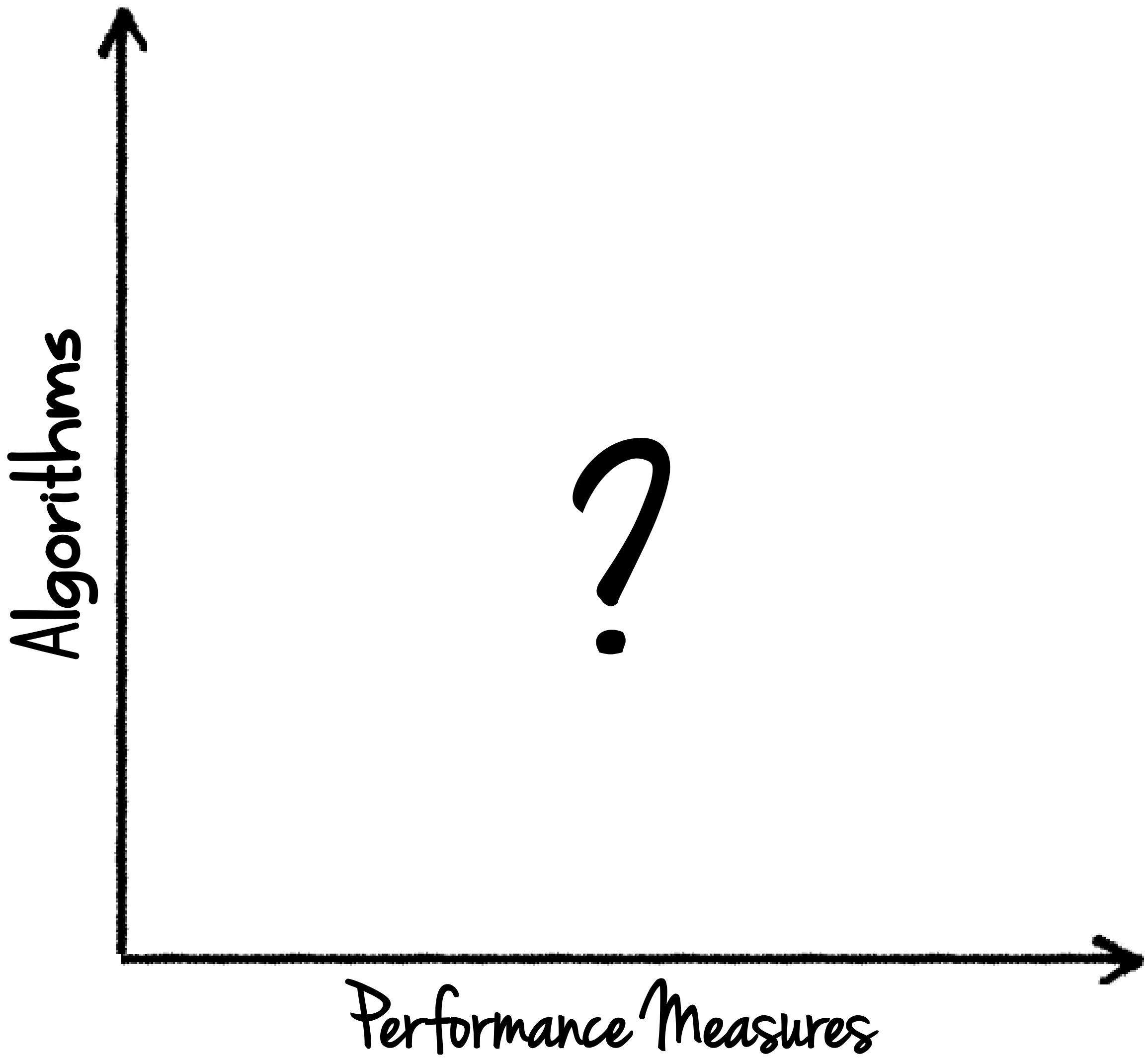
Partial Area Under the ROC Curve (AUC)



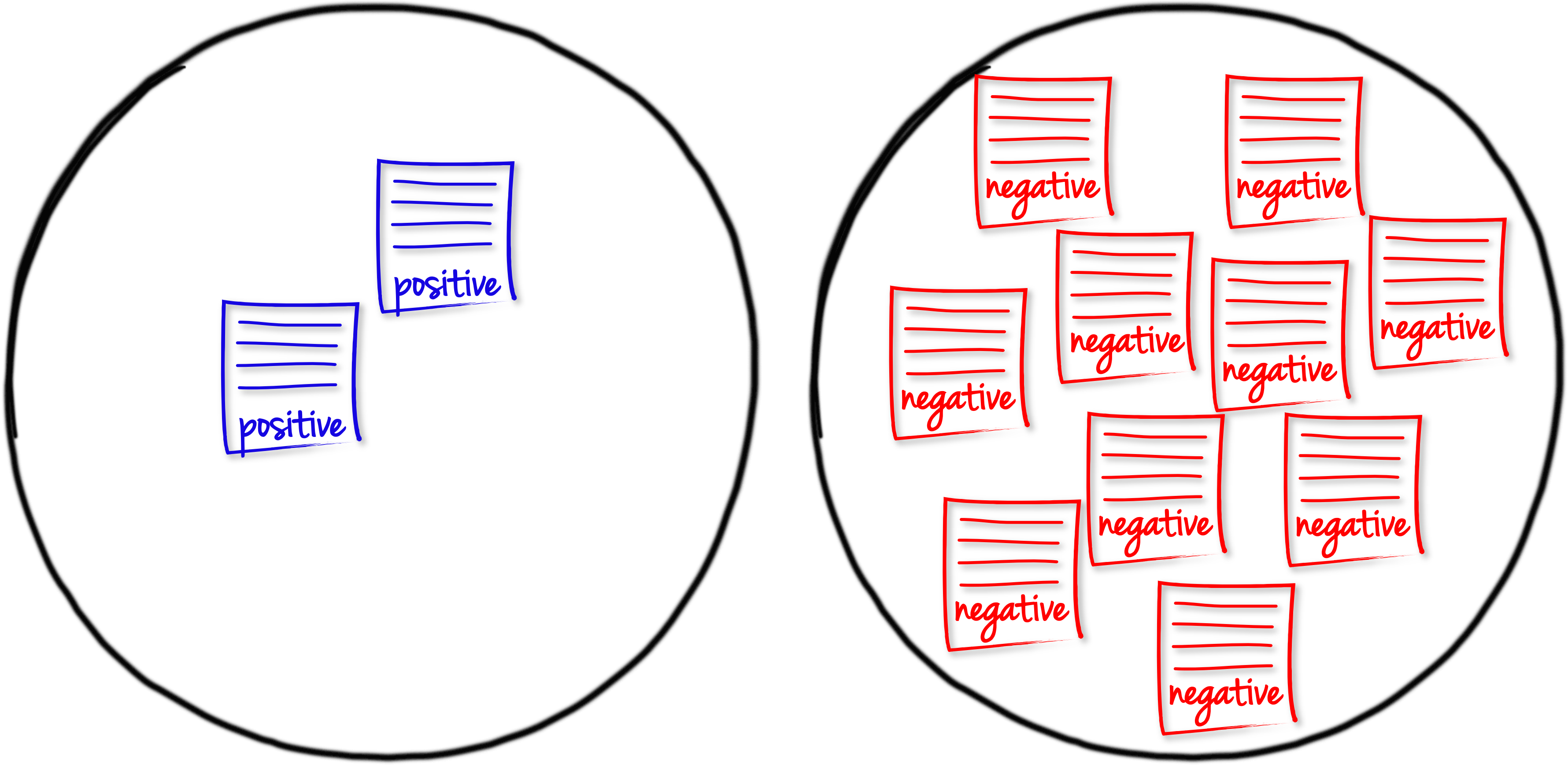
http://en.wikipedia.org/



http://www.tagxedo.com

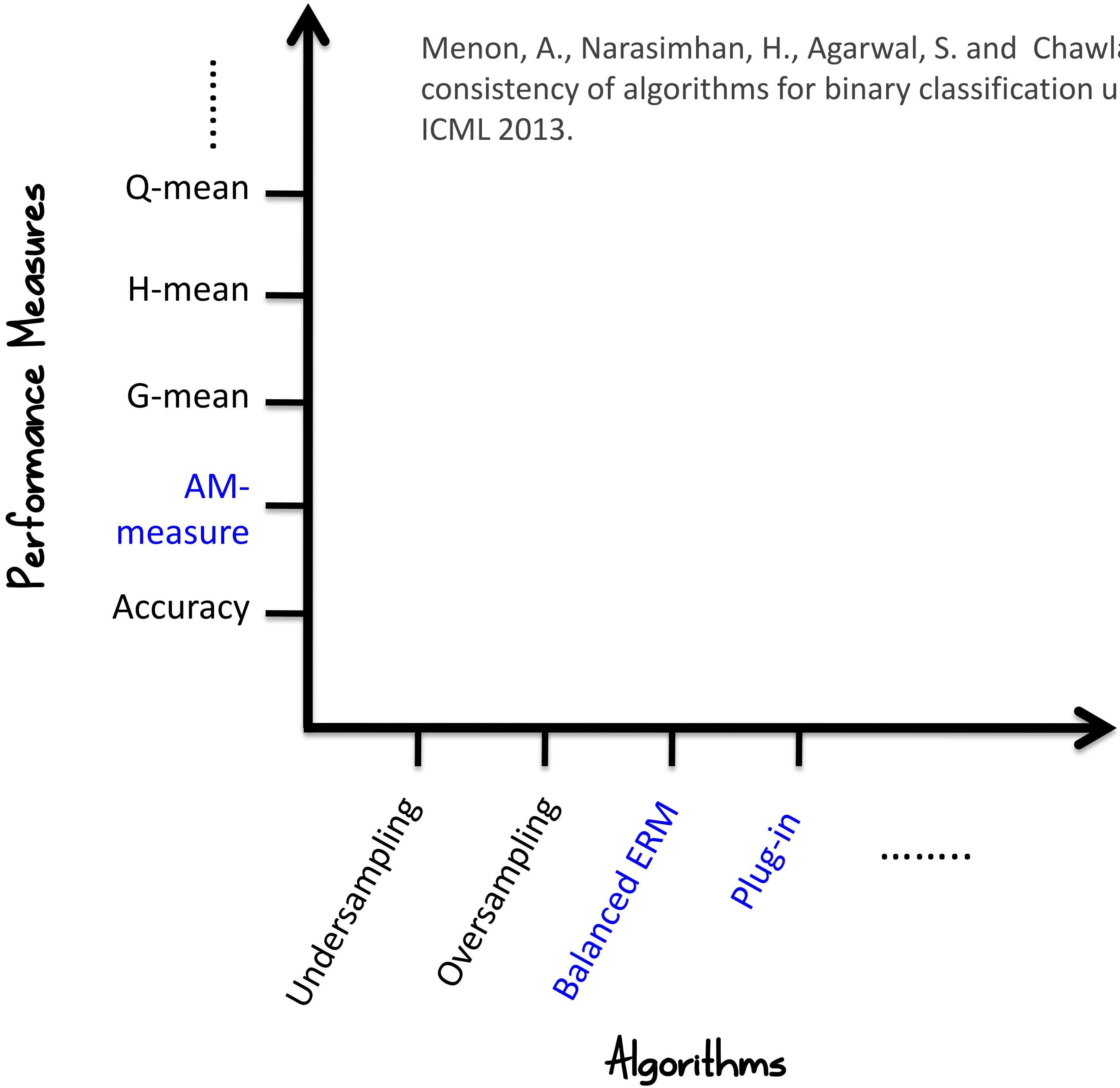

Algorithms
?
Performance Measures

Example: Class Imbalanced

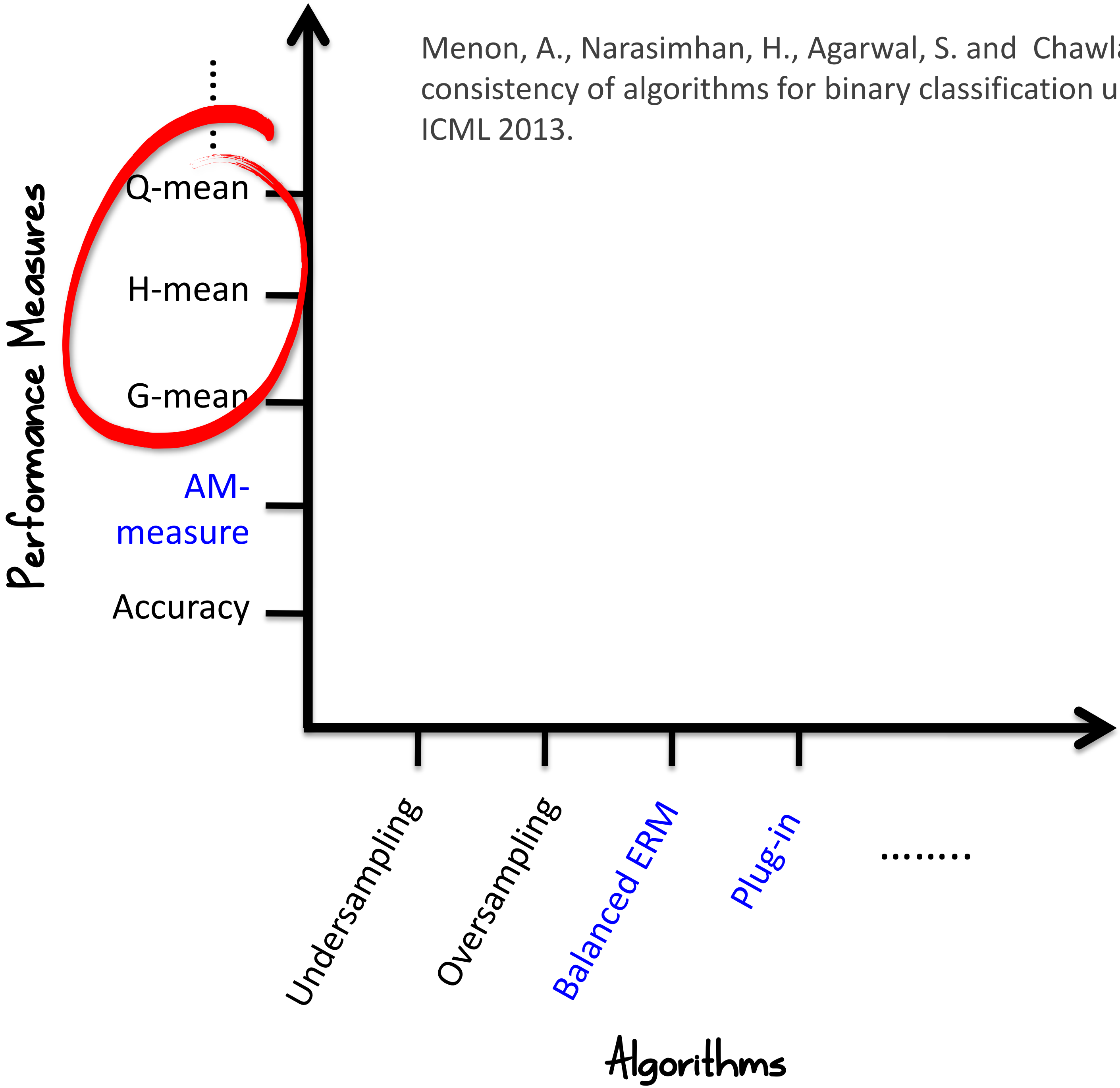


Standard Classification Error Ill-suited!

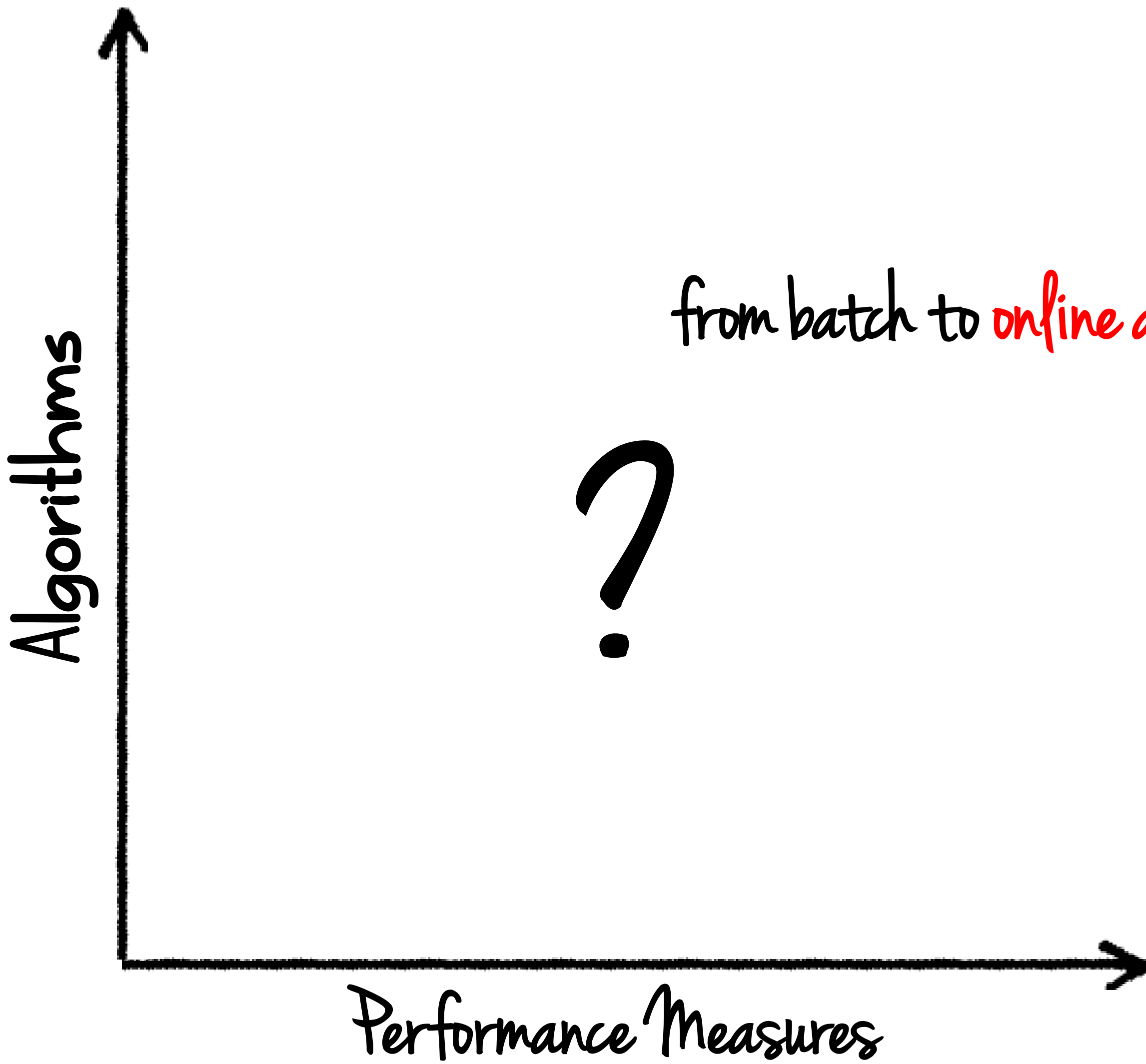
Menon, A., Narasimhan, H., Agarwal, S. and Chawla, S. “On the statistical consistency of algorithms for binary classification under class imbalance ”, ICML 2013.



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** . . .

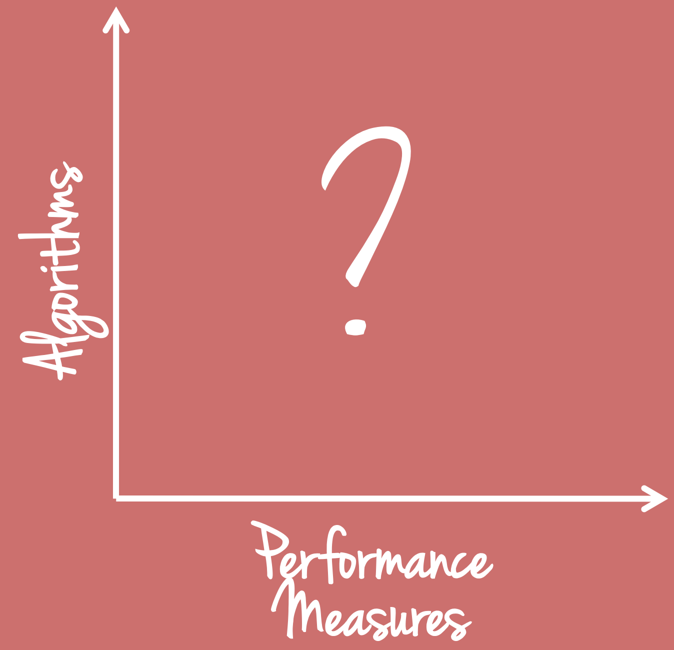


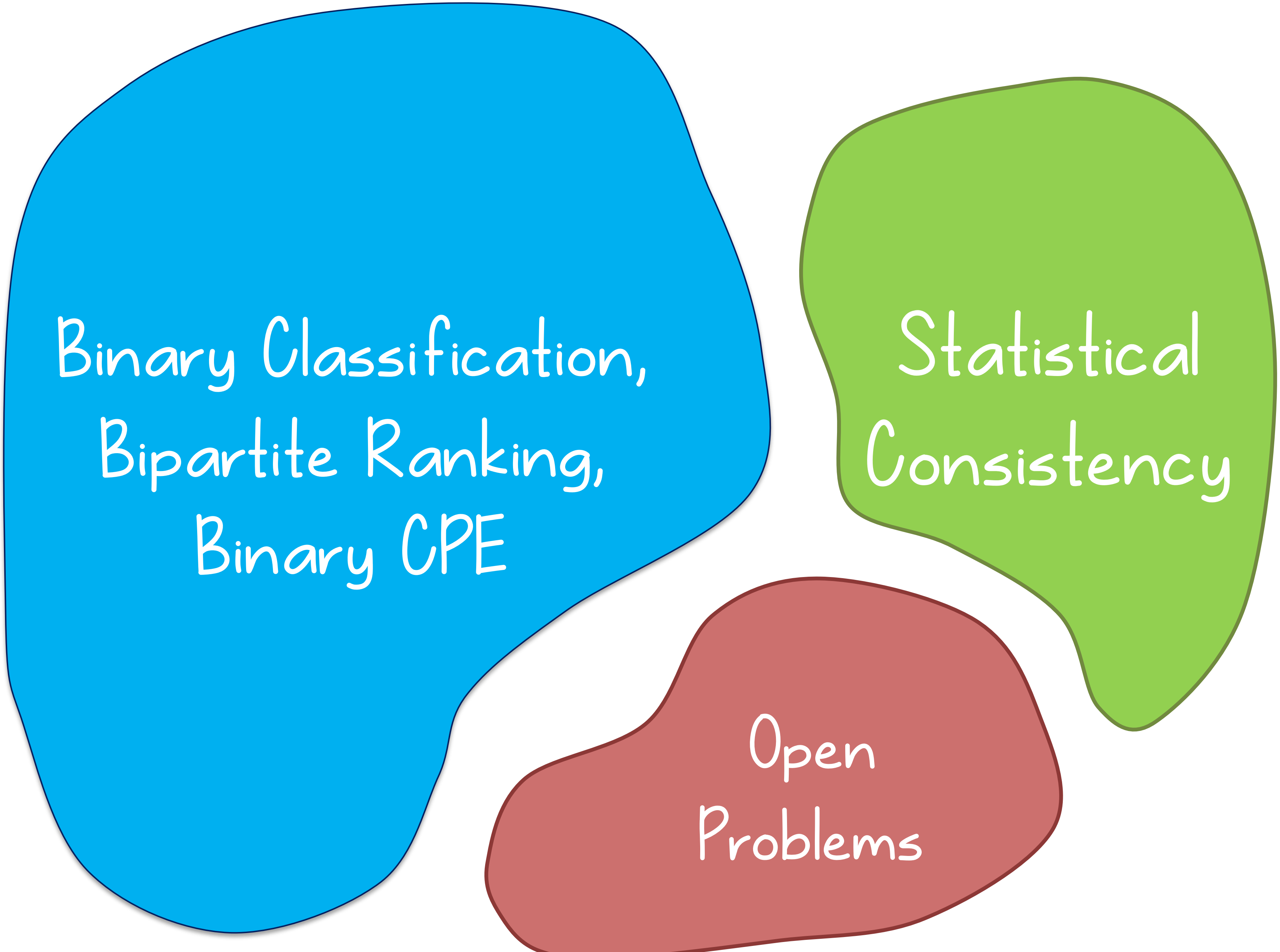
Batch Algorithms for Multivariate Performance Measures
(Structural SVM)

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?