



Kno.E.SIS

WRIGHT STATE
UNIVERSITY

COLLECTING THE DOTS | CONNECTING THE DOTS

Minimum Variance Semi-Supervised Boosting for Multi-label Classification

Chenyang Zhao, Shaodan Zhai



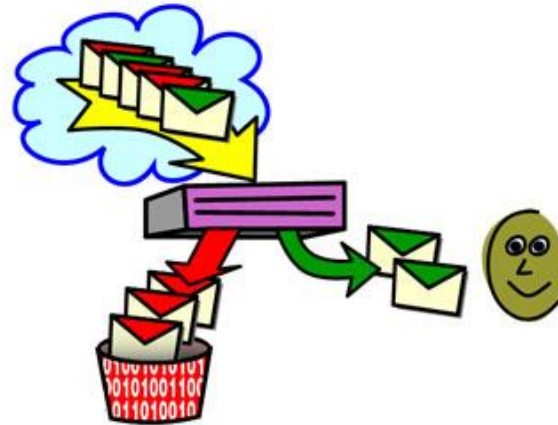
WRIGHT STATE
UNIVERSITY

Classification

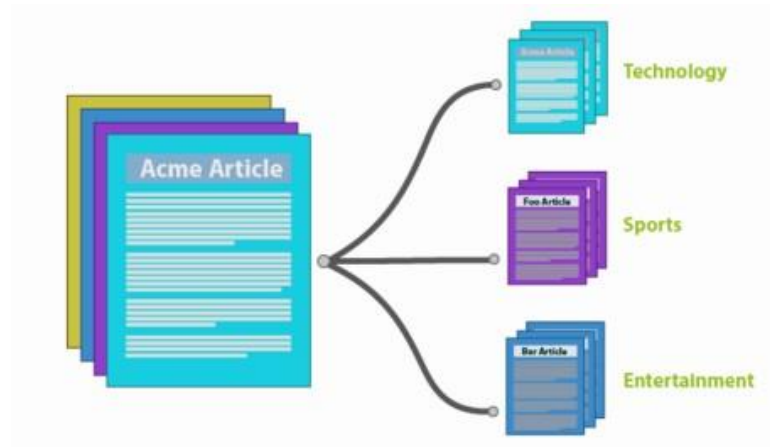
- Predicting a target label for a given instance

- For examples:

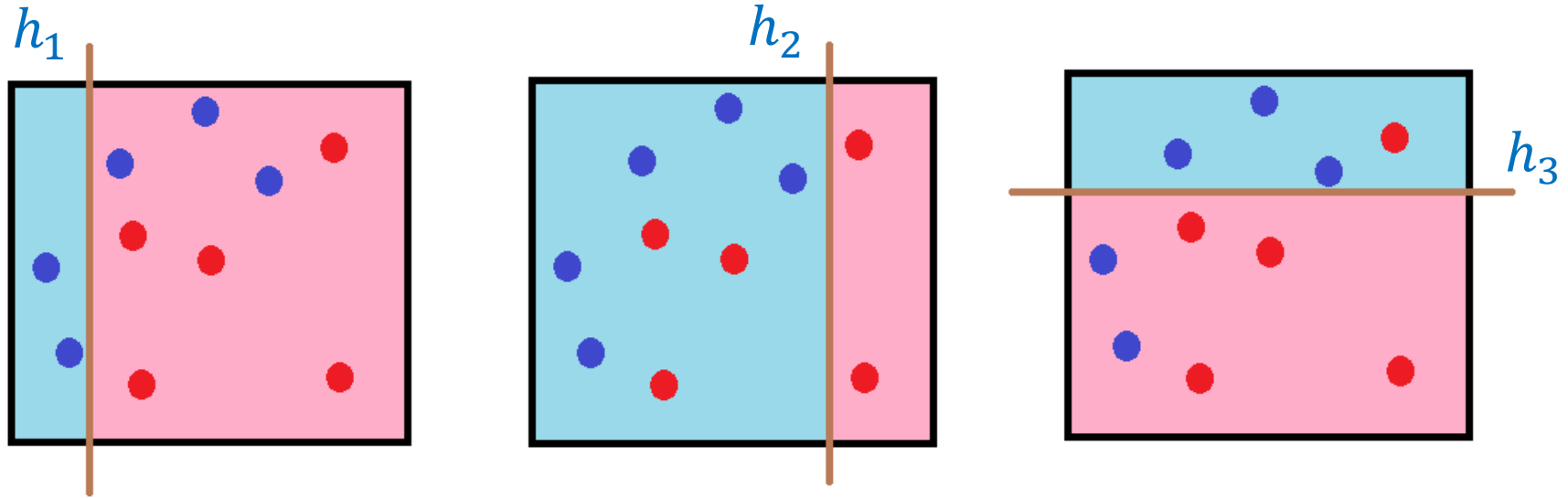
- Spam-detection



- Document categorization

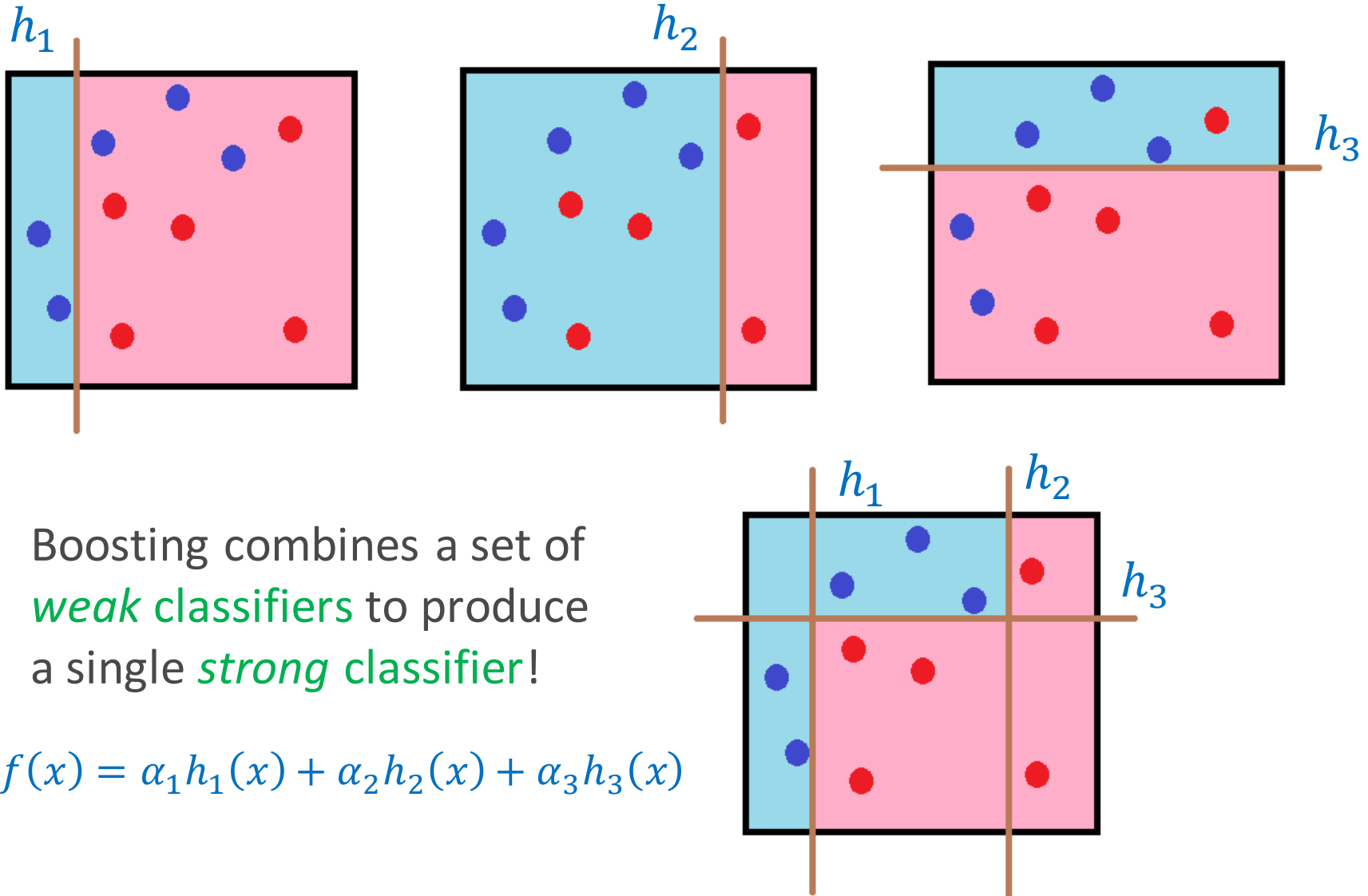


Boosting



Boosting combines a set of *weak classifiers* to produce a single *strong classifier*!

Boosting

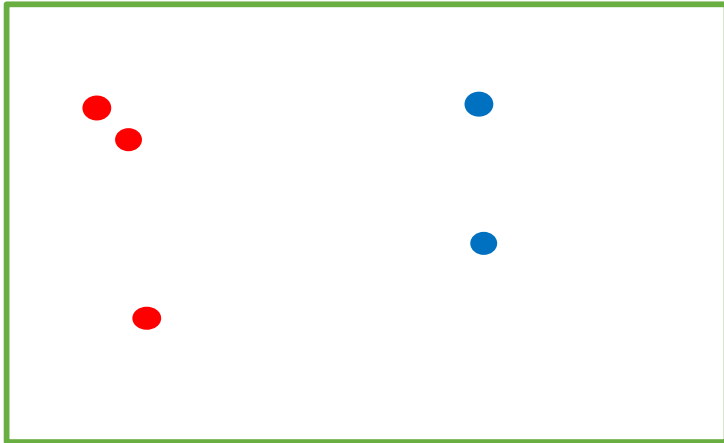


Boosting combines a set of *weak classifiers* to produce a single *strong classifier*!

$$f(x) = \alpha_1 h_1(x) + \alpha_2 h_2(x) + \alpha_3 h_3(x)$$

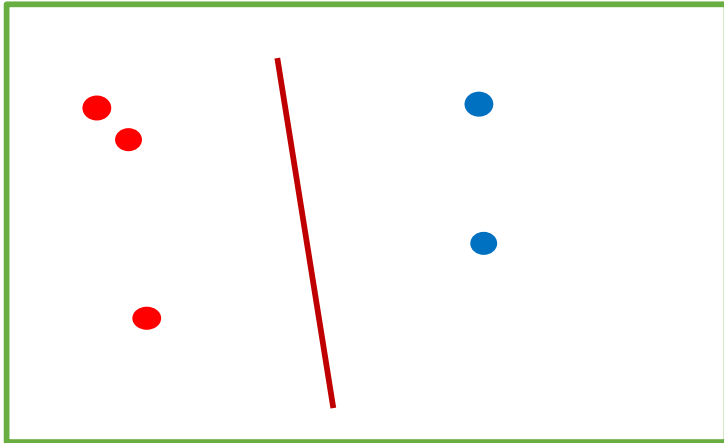
Semi-supervised Classification

- supervised learning use **labeled** training data



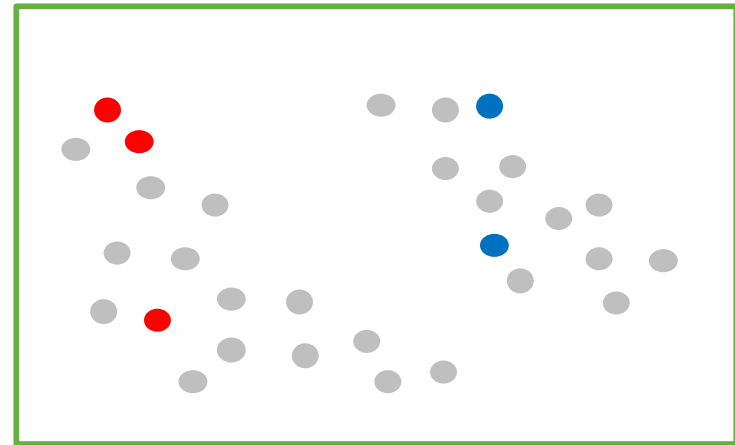
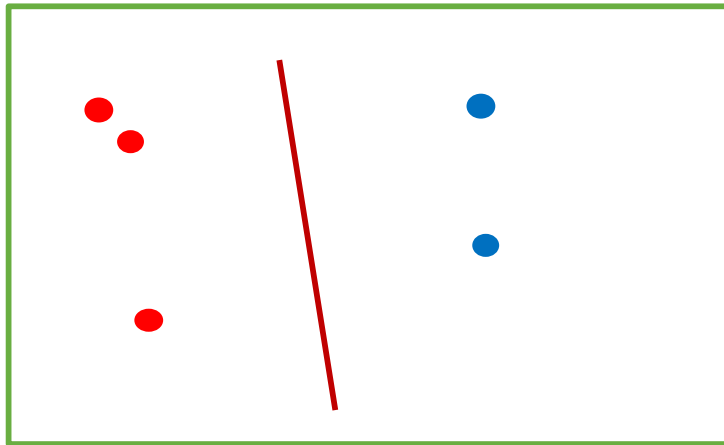
Semi-supervised Classification

- supervised learning use **labeled** training data



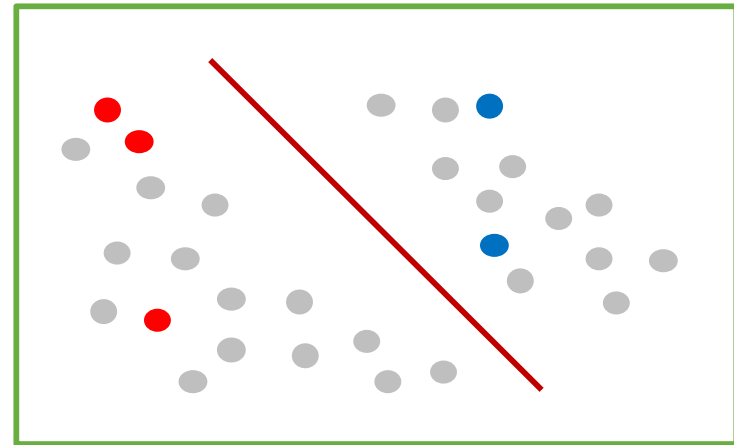
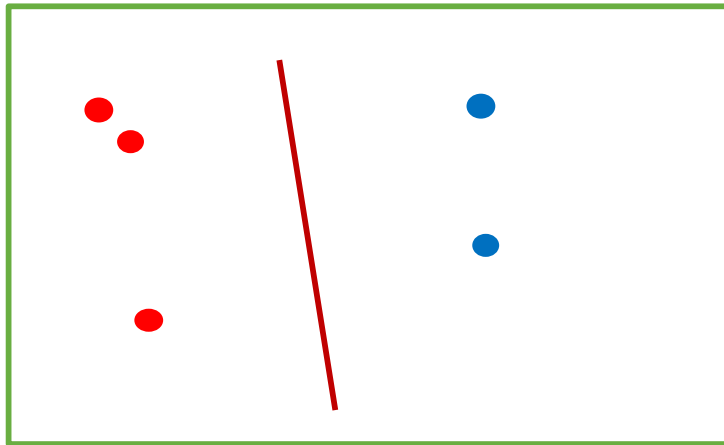
Semi-supervised Classification

- supervised learning use **labeled** training data
- can the classifier be improved by **unlabeled** data?



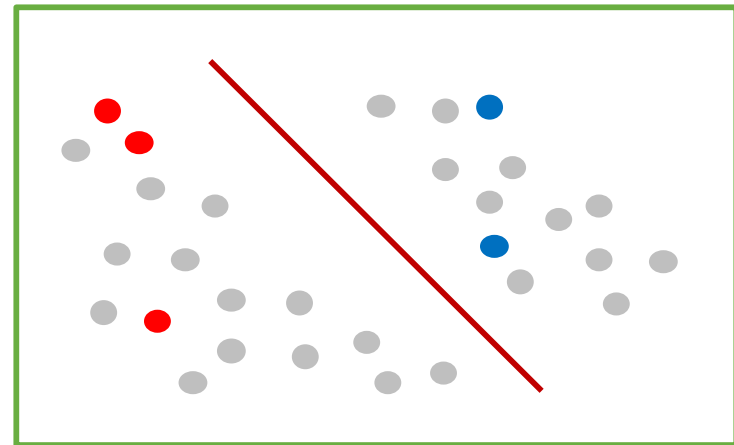
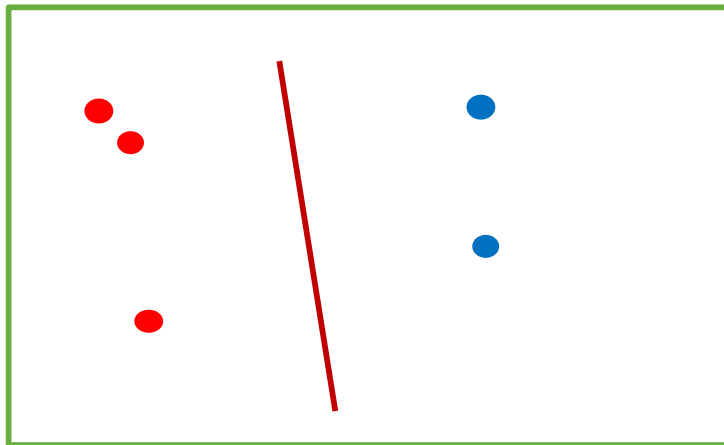
Semi-supervised Classification

- supervised learning use **labeled** training data
- can the classifier be improved by **unlabeled** data?



Semi-supervised Classification

- supervised learning use **labeled** training data
- can the classifier be improved by **unlabeled** data?



- when semi-supervised learning works?
 - labeled data is **limited** and **expensive**
 - unlabeled data is **plentiful** and **cheap**

Multi-label Classification

- one instance can belong to **more than one** category!

Multi-label Classification

- one instance can belong to **more than one** category!

LeBron James meets Prince William and Duchess Kate after Cavaliers-Nets

By *Seth Rosenthal* @seth_rosenthal on Dec 8, 2014, 11:47p



Neilson Barnard/Getty Images

TWEET (43)

SHARE (1220)

PIN

After the royal couple **watched the Cavs beat the Nets and fraternized with fellow Illuminati members** in Brooklyn, they met LeBron James for some photos!

They took some very nice pictures!



LATEST NEWS

SB Nation's 2015 NBA season preview

What we learned from the Tristan Thompson saga

Olympic skier Gus Kenworthy comes out as gay

Dodgers announce they've parted ways with Mattingly

POWERED BY **IHOP**

Watch Kids In Adorable Costumes Being Adorable While Playing With IHOP's Scary Face Pancakes

Sticking with Cousins is costing Washington wine

Multi-label Classification

- one instance can belong to **more than one** category!

LeBron James meets Prince William and Duchess Kate after Cavaliers-Nets

By *Seth Rosenthal* @seth_roenthal on Dec 8, 2014, 11:47p



Neilson Barnard/Getty Images

TWEET (43)

SHARE (1220)

PIN

After the royal couple **watched the Cavs beat the Nets and fraternized with fellow Illuminati members** in Brooklyn, they met LeBron James for some photos!

They took some very nice pictures!



Sports

LATEST NEWS

SB Nation's 2015 NBA season preview

What we learned from the Tristan Thompson saga

Olympic skier Gus Kenworthy comes out as gay

Dodgers announce they've parted ways with Mattingly

POWERED BY **IHOP**

Watch Kids In Adorable Costumes Being Adorable While Playing With IHOP's Scary Face Pancakes

Sticking with Cousins is costing Washington wine

Multi-label Classification

- one instance can belong to **more than one** category!

LeBron James meets Prince William and Duchess Kate after Cavaliers-Nets

By *Seth Rosenthal* @seth_rosenthal on Dec 8, 2014, 11:47p



Neilson Barnard/Getty Images

TWEET (43)

SHARE (1220)

PIN

After the royal couple **watched the Cavs beat the Nets and fraternized with fellow Illuminati members** in Brooklyn, they met LeBron James for some photos!

They took some very nice pictures!



LATEST NEWS

SB Nation's 2015 NBA season preview

What we learned from the Tristan Thompson saga

Olympic skier Gus Kenworthy comes out as gay

Dodgers announce they've parted ways with Mattingly

POWERED BY **IHOP**

Watch Kids In Adorable Costumes Being Adorable While Playing With IHOP's Scary Face Pancakes


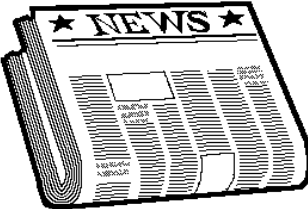

Sticking with Cousins is costing Washington wine

Sports

Politics

Multi-label Classification

- one instance can belong to **more than one** category!
- a learning task of predicting **a set of target labels** for a instance

	Politics	Economy	Sports	Business	Art
	✓		✓		
		✓		✓	
					✓

Generic Semi-supervised Multi-label Boosting Algorithm

- given n labeled data $(x_1, Y_1), \dots, (x_n, Y_n)$ and m unlabeled data x_{n+1}, \dots, x_{n+m}

Generic Semi-supervised Multi-label Boosting Algorithm

- given n labeled data $(x_1, Y_1), \dots, (x_n, Y_n)$ and m unlabeled data x_{n+1}, \dots, x_{n+m}
- ensemble classifier for each label l

$$F_T^l = \alpha_1^l h_1^l(x) + \dots + \alpha_T^l h_T^l(x)$$

Generic Semi-supervised Multi-label Boosting Algorithm

- given n labeled data $(x_1, Y_1), \dots, (x_n, Y_n)$ and m unlabeled data x_{n+1}, \dots, x_{n+m}

- ensemble classifier for each label l

$$F_T^l = \alpha_1^l h_1^l(x) + \dots + \alpha_T^l h_T^l(x)$$

- minimize a loss function of on F both labeled and unlabeled data

$$J(F) = \sum_{i=1}^m \sum_{l=1}^K J_L(F^l(x_i)) + \gamma \sum_{i=m+1}^m \sum_{l=1}^K J_U(F^l(x_i))$$

Generic Semi-supervised Multi-label Boosting Algorithm

- given n labeled data $(x_1, Y_1), \dots, (x_n, Y_n)$ and m unlabeled data x_{n+1}, \dots, x_{n+m}

- ensemble classifier for each label l

$$F_T^l = \alpha_1^l h_1^l(x) + \dots + \alpha_T^l h_T^l(x)$$

- minimize a loss function of on F both labeled and unlabeled data

$$J(F) = \sum_{i=1}^m \sum_{l=1}^K J_L(F^l(x_i)) + \gamma \sum_{i=m+1}^m \sum_{l=1}^K J_U(F^l(x_i))$$

labeled margin

Generic Semi-supervised Multi-label Boosting Algorithm

- given n labeled data $(x_1, Y_1), \dots, (x_n, Y_n)$ and m unlabeled data x_{n+1}, \dots, x_{n+m}

- ensemble classifier for each label l

$$F_T^l = \alpha_1^l h_1^l(x) + \dots + \alpha_T^l h_T^l(x)$$

- minimize a loss function of on F both labeled and unlabeled data

$$J(F) = \sum_{i=1}^m \sum_{l=1}^K J_L(F^l(x_i)) + \gamma \sum_{i=m+1}^m \sum_{l=1}^K J_U(F^l(x_i))$$

unlabeled margin

Generic Semi-supervised Multi-label Boosting Algorithm

- given n labeled data $(x_1, Y_1), \dots, (x_n, Y_n)$ and m unlabeled data x_{n+1}, \dots, x_{n+m}

- ensemble classifier for each label l

$$F_T^l = \alpha_1^l h_1^l(x) + \dots + \alpha_T^l h_T^l(x)$$

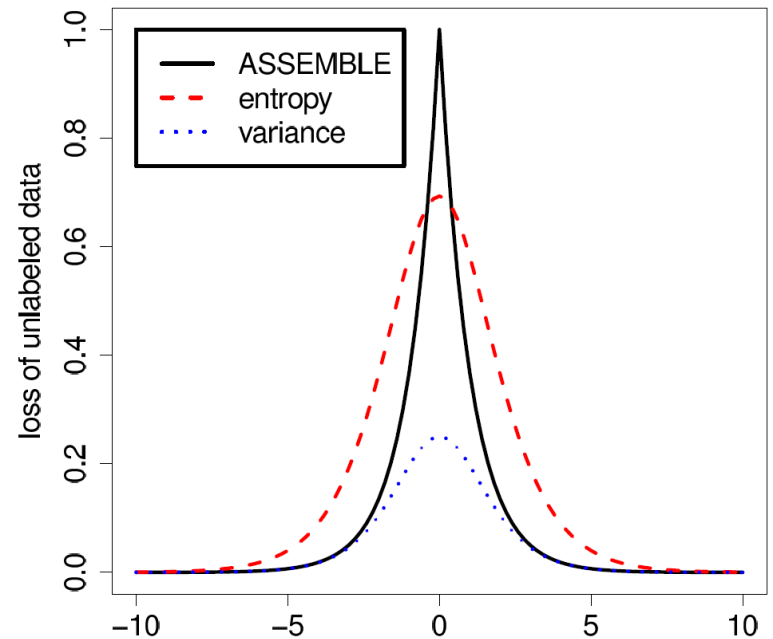
- minimize a loss function of on F both labeled and unlabeled data

$$J(F) = \sum_{i=1}^m \sum_{l=1}^K J_L(F^l(x_i)) + \gamma \sum_{i=m+1}^m \sum_{l=1}^K J_U(F^l(x_i))$$

trade-off parameter

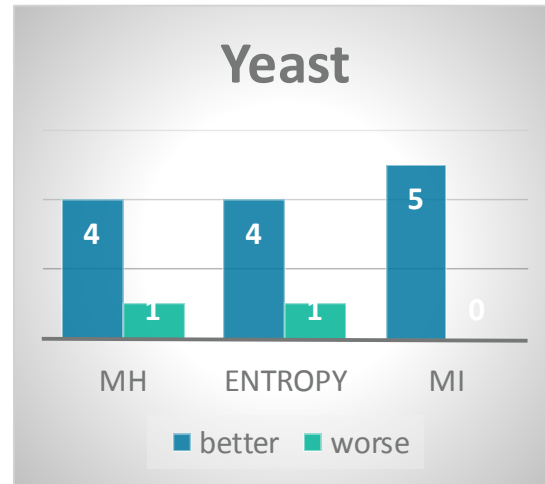
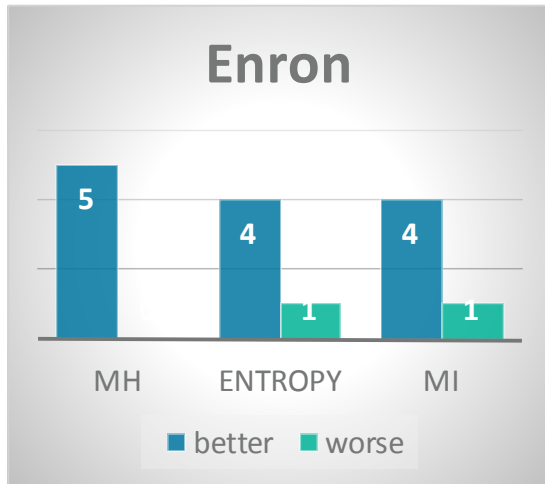
Minimizing Variance Regularization over Unlabeled Data

- to use **conditional variance** $Var(Y|x)$ as a loss function over unlabeled data
 - Encouraging F to be large in magnitude
 - Variance expresses measurement uncertainty of the label variable
 - Can be viewed as minimizing the expected value of a sigmoid loss
 - Smooth, differentiable function of $F(x)$



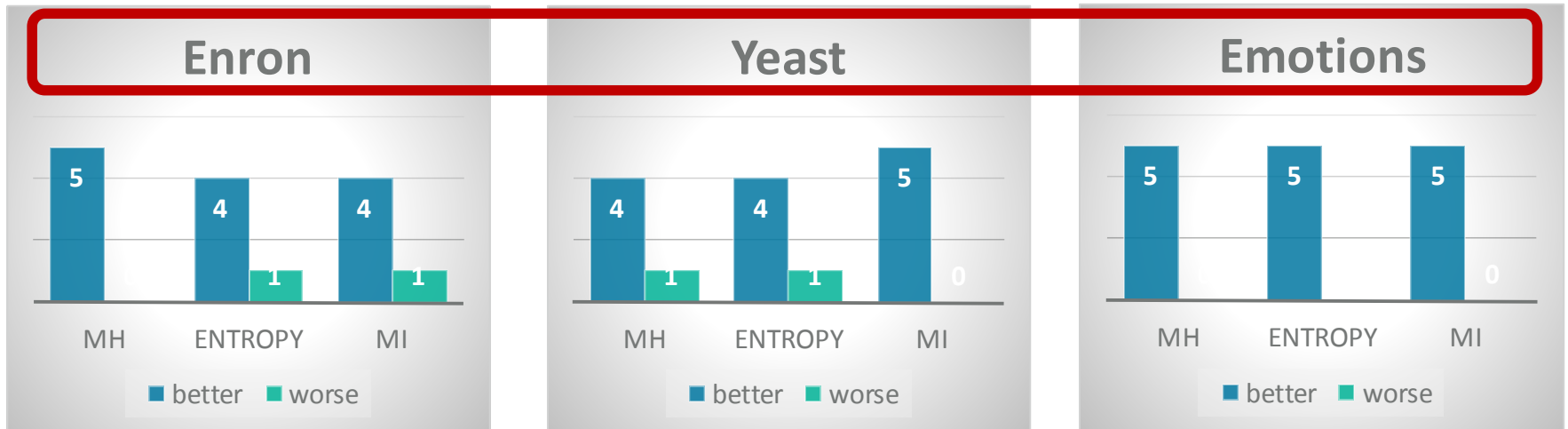
Experiments

Measured on **Hamming loss**, **Accuracy**, **Precision**, **Recall** and **F-value**



Experiments

Measured on **Hamming loss**, **Accuracy**, **Precision**, **Recall** and **F-value**



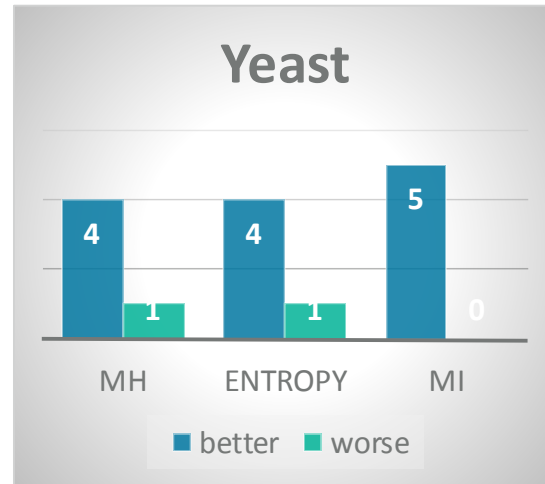
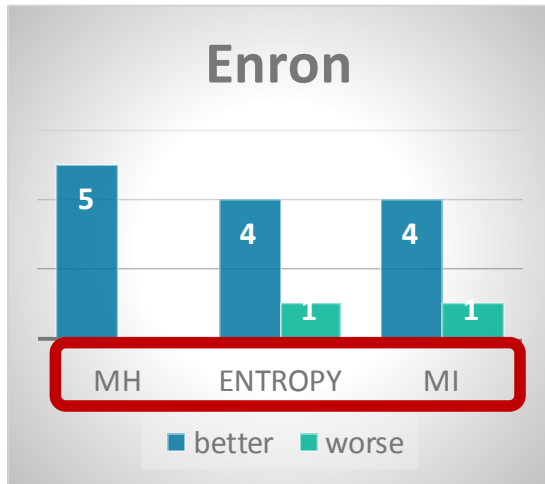
Enron: text categorization

Yeast: gene function classification

Emotions: emotion classification

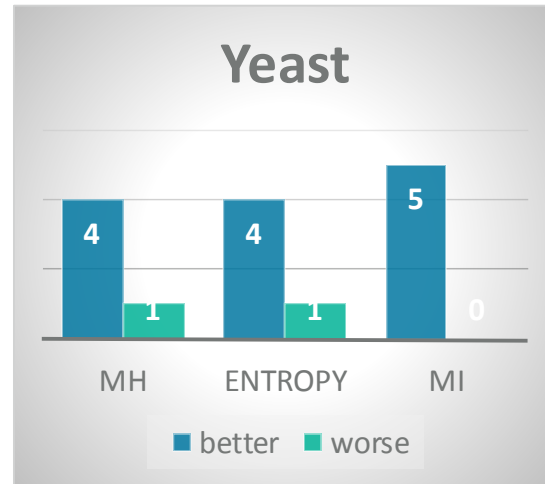
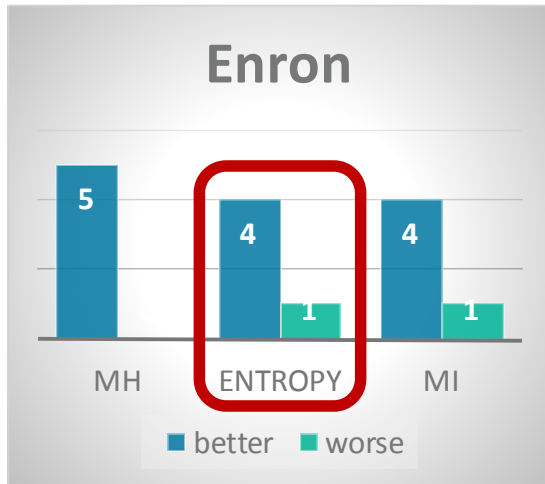
Experiments

Measured on **Hamming loss**, **Accuracy**, **Precision**, **Recall** and **F-value**



Experiments

Measured on **Hamming loss**, **Accuracy**, **Precision**, **Recall** and **F-value**



Thank You! & Questions?