

# Pessimism is Great for Machine Learning!

Brian Ziebart

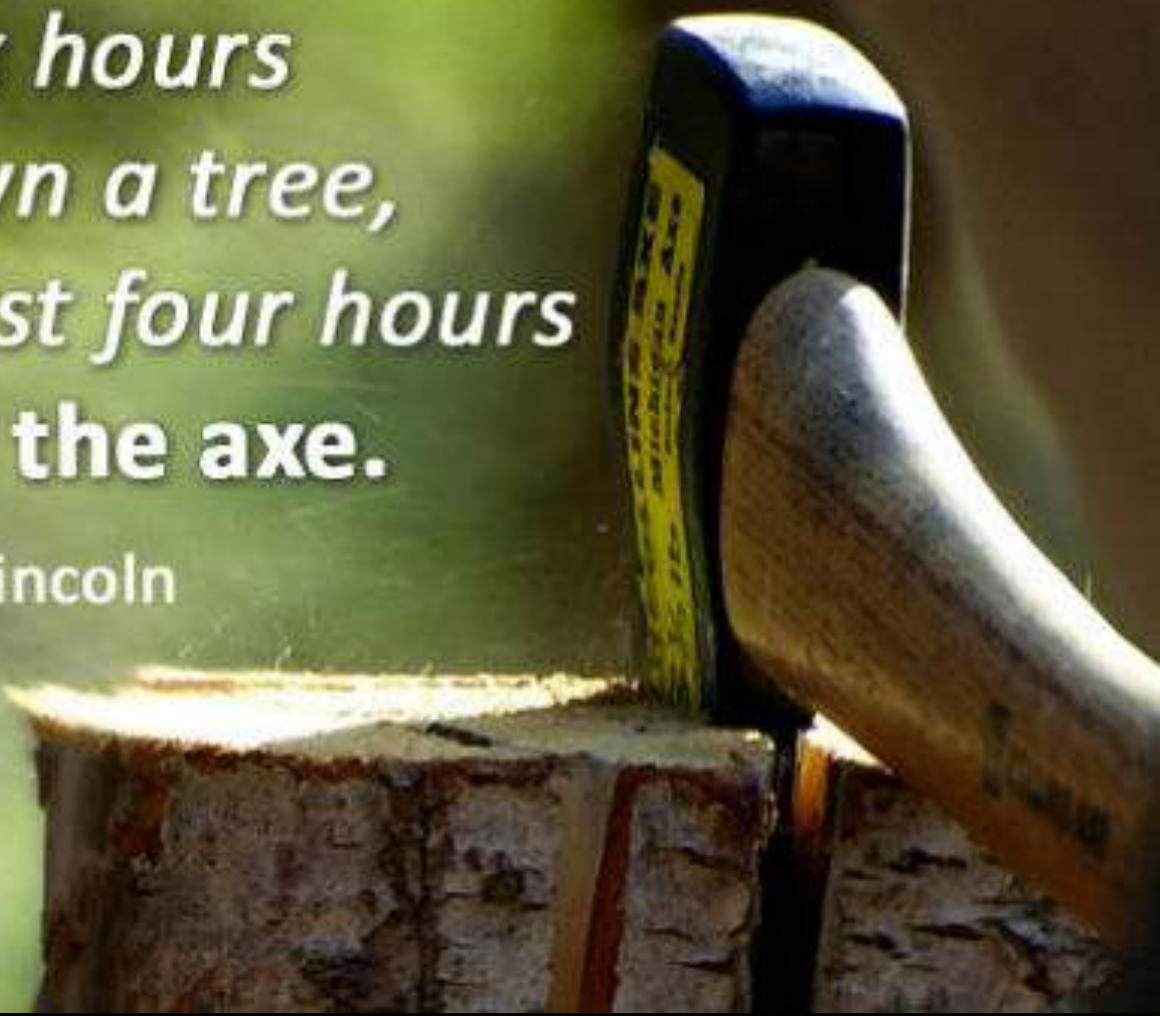
[bziebart@uic.edu](mailto:bziebart@uic.edu)



This project supports the PhD thesis work of Anqi Liu ([aliu33@uic.edu](mailto:aliu33@uic.edu))

*If I had six hours  
to chop down a tree,  
I'd spend the first four hours  
sharpening the axe.*

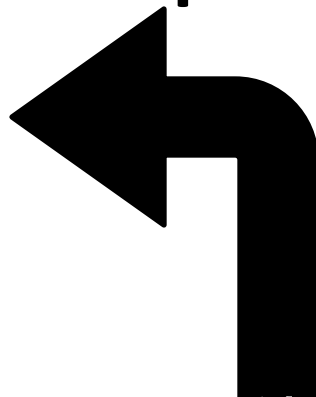
~ Abraham Lincoln



# Training



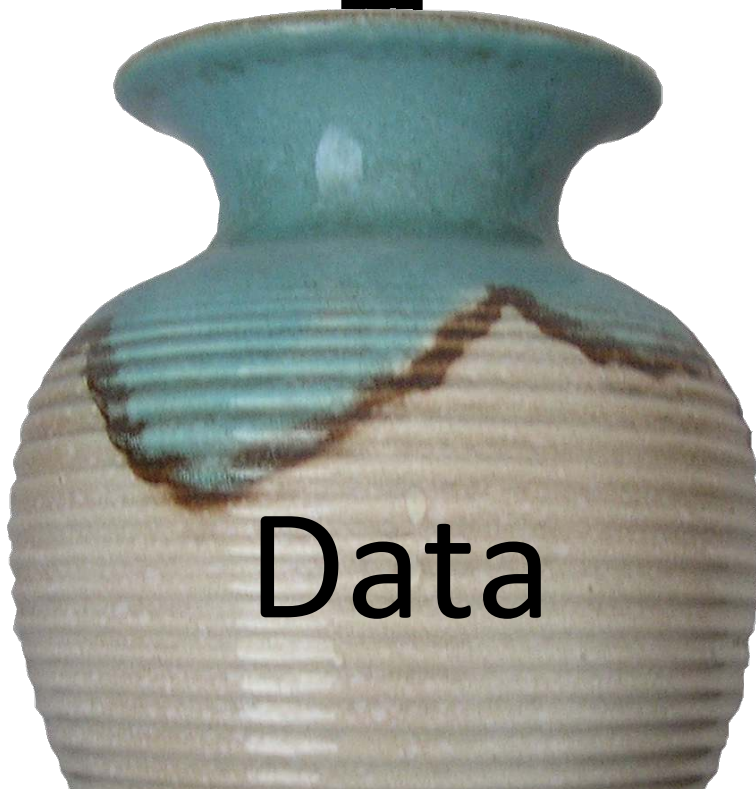
M samples



Sample dist.

$\tilde{P}(\mathbf{x}, \mathbf{y})$

Data



# Training



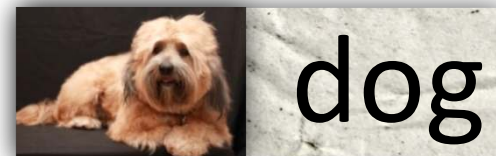
Sample dist.  
 $\tilde{P}(\mathbf{x}, \mathbf{y})$

dog

Data

# Testing

Prediction:  $\hat{y}(\mathbf{x})$



Loss:  $\text{loss}(\hat{y}, y)$



# Testing

Prediction:  $\hat{y}(x)$



Loss:  $\text{loss}(\hat{y}, y)$

Expected Loss:  
 $E_p[\text{loss}(\hat{y}(X), Y)]$

Sample dist.  
 $\tilde{P}(x, y)$

True dist.  
 $P(x, y)$

Data



|     | $y$ |     |     |
|-----|-----|-----|-----|
|     | Dog | Cat | Car |
| Dog | 0   | 1   | 1   |
| Cat | 1   | 0   | 1   |
| Car | 1   | 1   | 0   |

|     | $y$ |     |     |
|-----|-----|-----|-----|
|     | Dog | Cat | Car |
| Dog | 0   | 1   | 5   |
| Cat | 1   | 0   | 5   |
| Car | 5   | 5   | 0   |

# Empirical Risk Minimization

Minimize approximate loss on  
exact training data



# Adversarial Risk Minimization

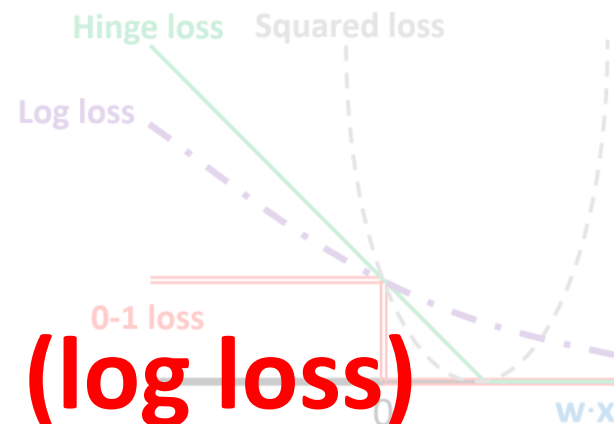
Minimize exact loss on  
approximate training data

|                        | $\mathbf{y}$ |     |     |
|------------------------|--------------|-----|-----|
|                        | Dog          | Cat | Car |
| $\hat{\mathbf{y}}$ Dog | 0            | 1   | 5   |
| Cat                    | 1            | 0   | 5   |
| Car                    | 5            | 5   | 0   |

Find equilibrium  $\hat{\mathbf{p}}, \mathbf{p}^*$

# Empirical Risk Minimization

Minimize approximate loss on  
exact training data



**Sometimes equivalent (log loss)**

**Pessimism is safer/better in general**

## Adversarial Risk Minimization

Minimize exact loss on  
approximate training data

zation

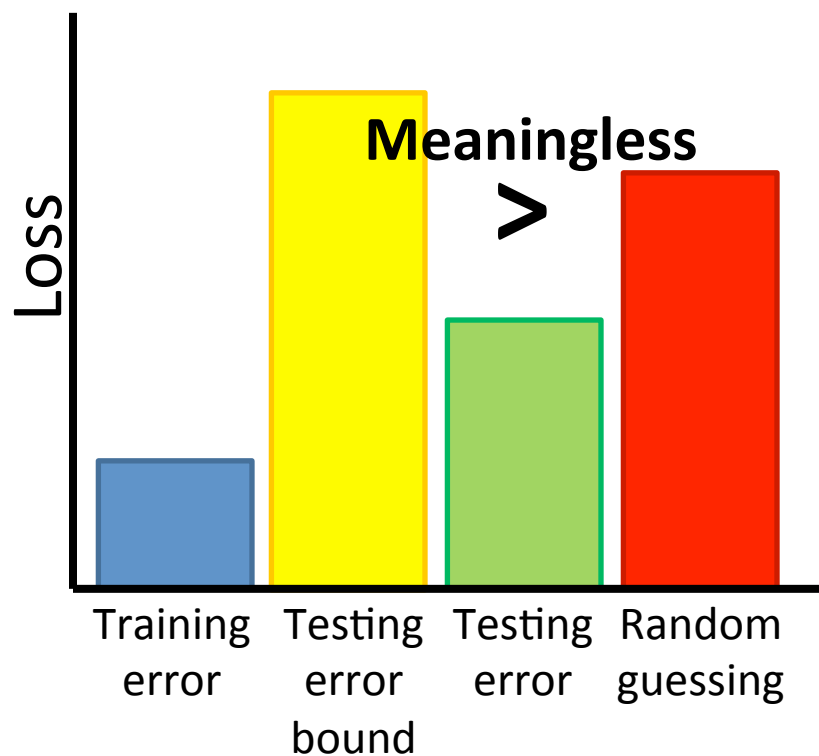
$\mathbf{y}$

|                        | Dog                   | Cat                   | Car                   |
|------------------------|-----------------------|-----------------------|-----------------------|
| $\hat{\mathbf{y}}$ Dog | $0+\psi_{\text{dog}}$ | $1+\psi_{\text{cat}}$ | $5+\psi_{\text{car}}$ |
| Cat                    | $1+\psi_{\text{dog}}$ | $0+\psi_{\text{cat}}$ | $5+\psi_{\text{car}}$ |
| Car                    | $5+\psi_{\text{dog}}$ | $5+\psi_{\text{cat}}$ | $0+\psi_{\text{car}}$ |

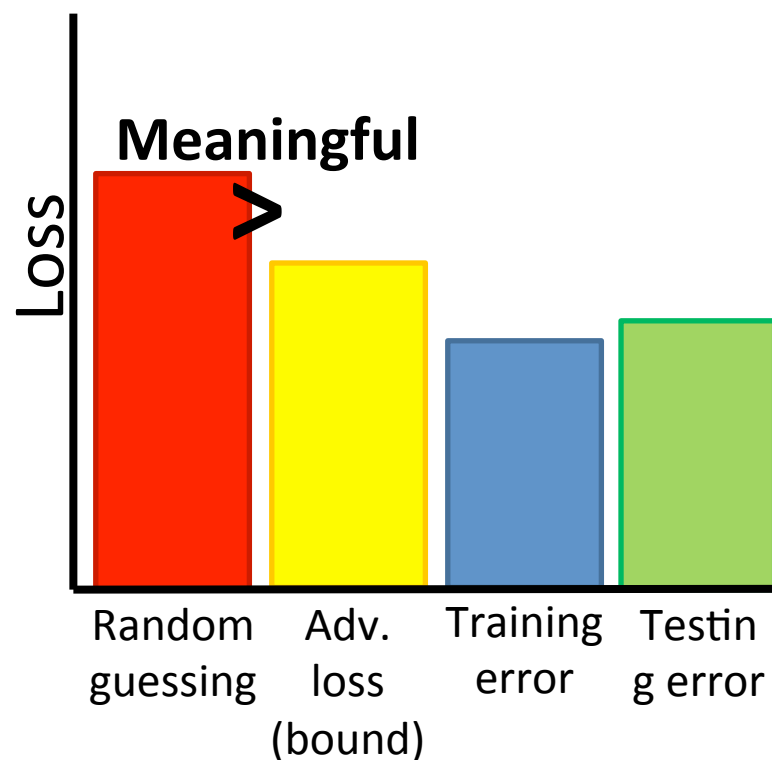
Find equilibrium  $\hat{\mathbf{p}}, \mathbf{p}$   
 $\psi_{\text{dog}} = \mathbf{w}_{\text{dog}} \cdot \mathbf{f}(\mathbf{x})$

# Meaningful Generalization Bounds

Empirical Risk Minimization



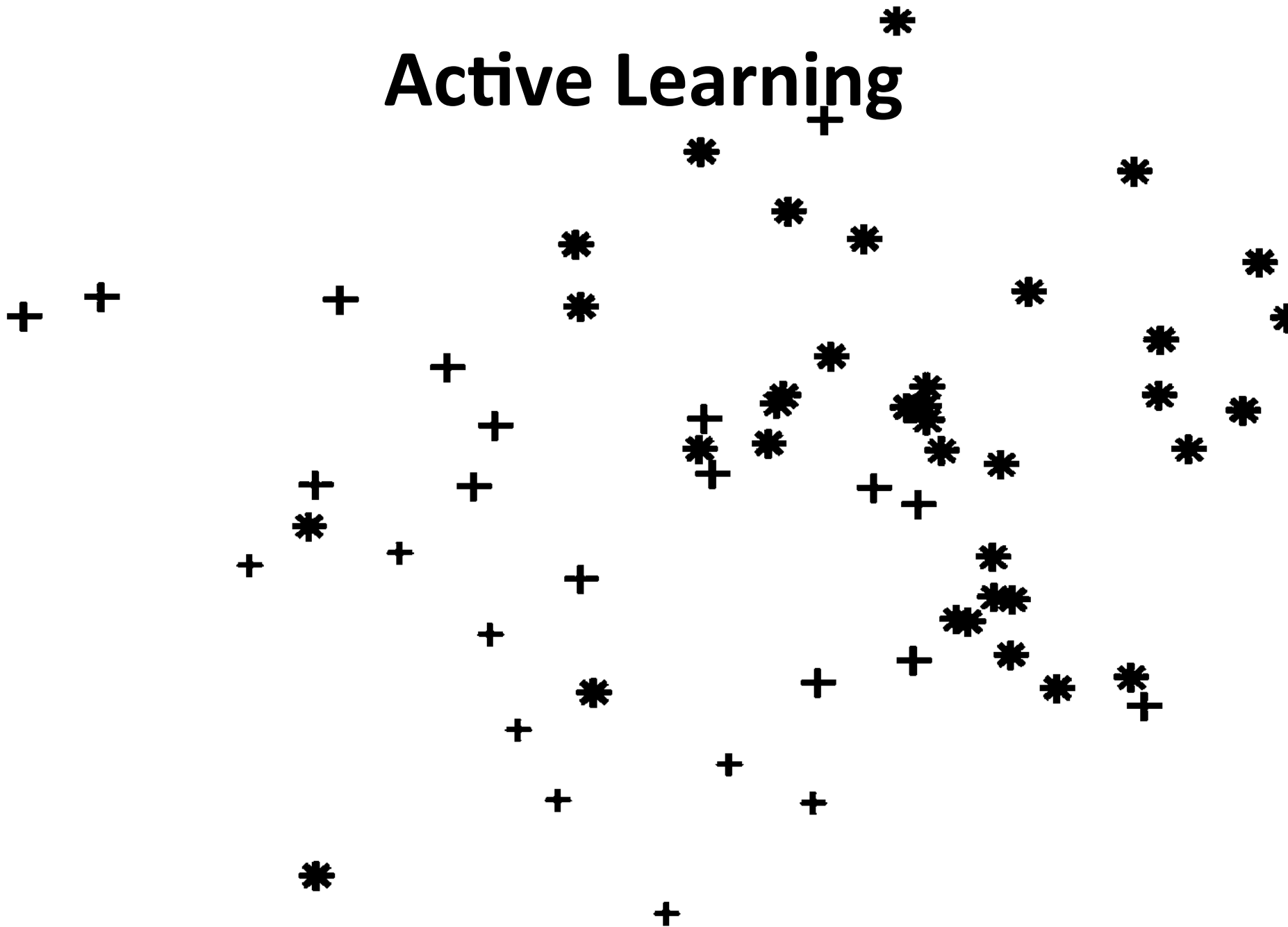
Adversarial Risk Minimization



Especially important when training and testing distributions differ (covariate shift) [Liu & Ziebart 2014, Liu et al. 2015, Chen et al. 2016]



# Active Learning



# ERM Active Learning

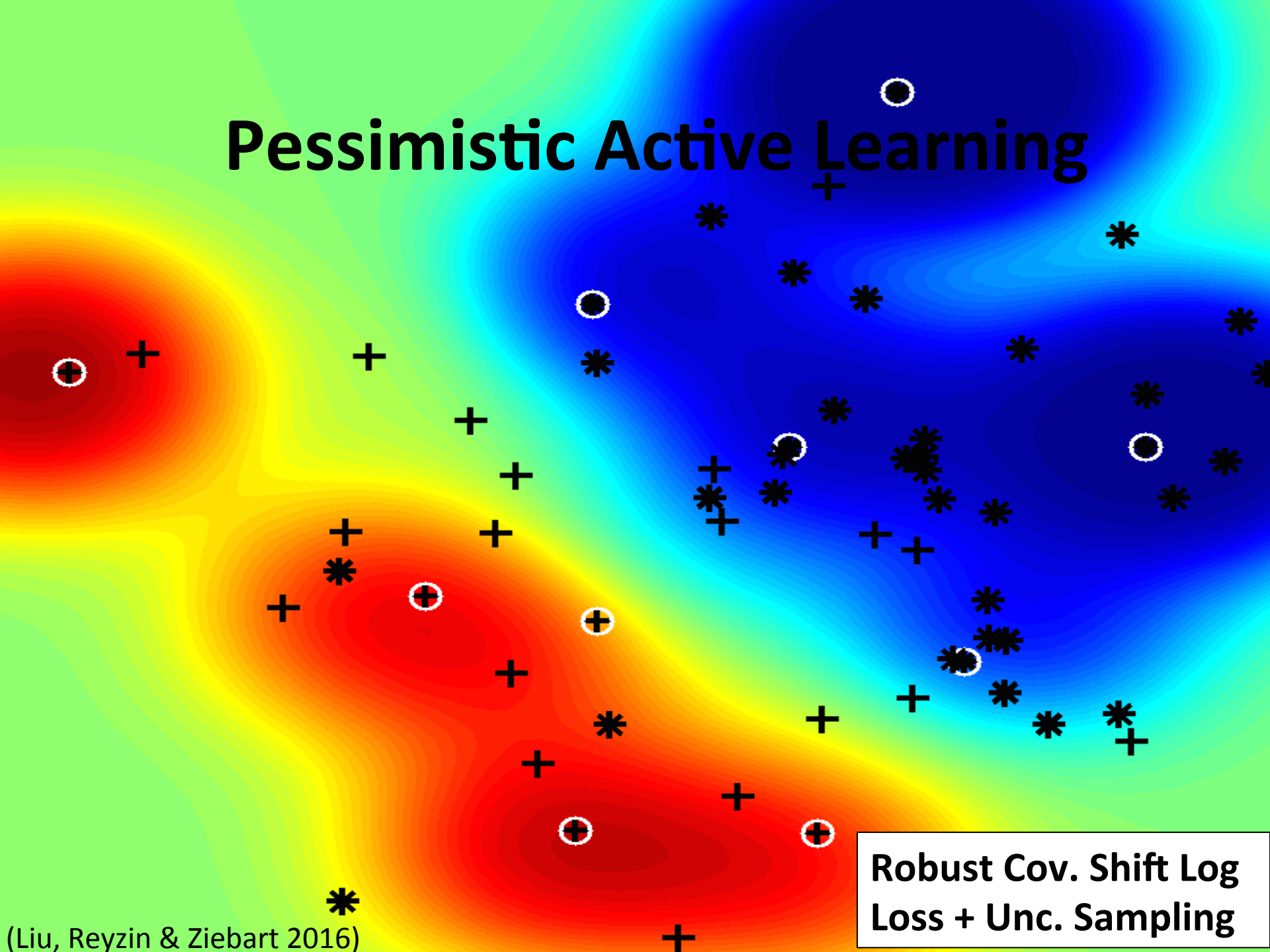
Wrong, but certain  
about datapoints  
over here

Labels all datapoints here  
before correcting the model

Good performance  
on labeled data

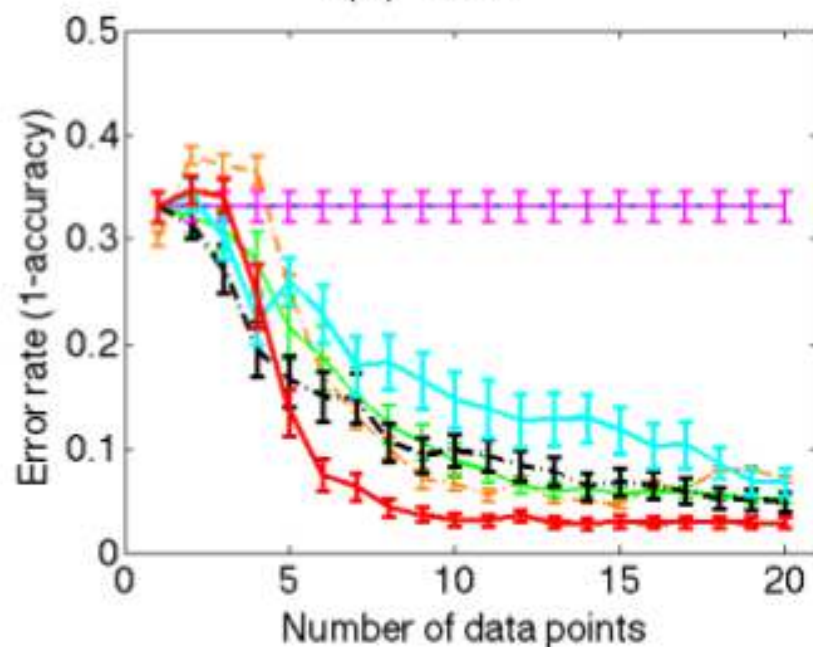
Logistic Regression +  
Uncertainty Sampling

# Pessimistic Active Learning

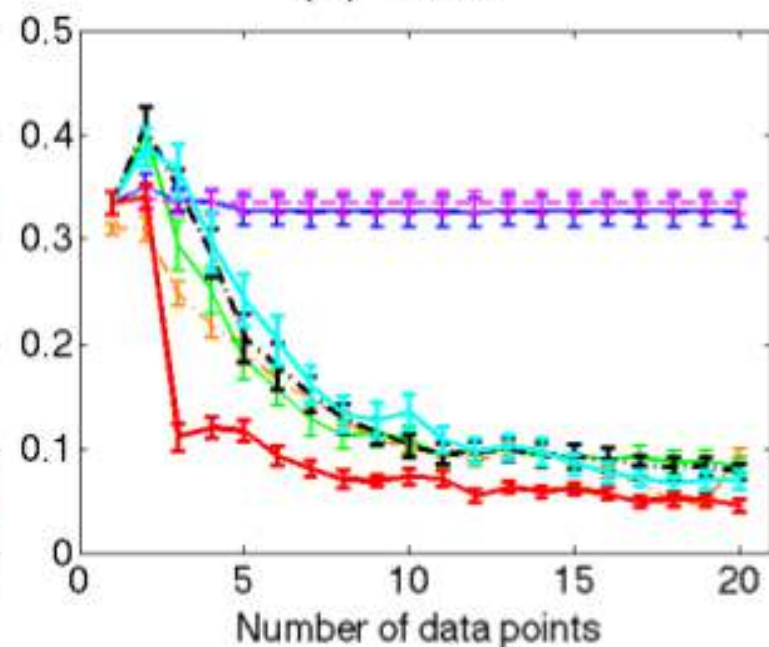


Robust Cov. Shift Log  
Loss + Unc. Sampling

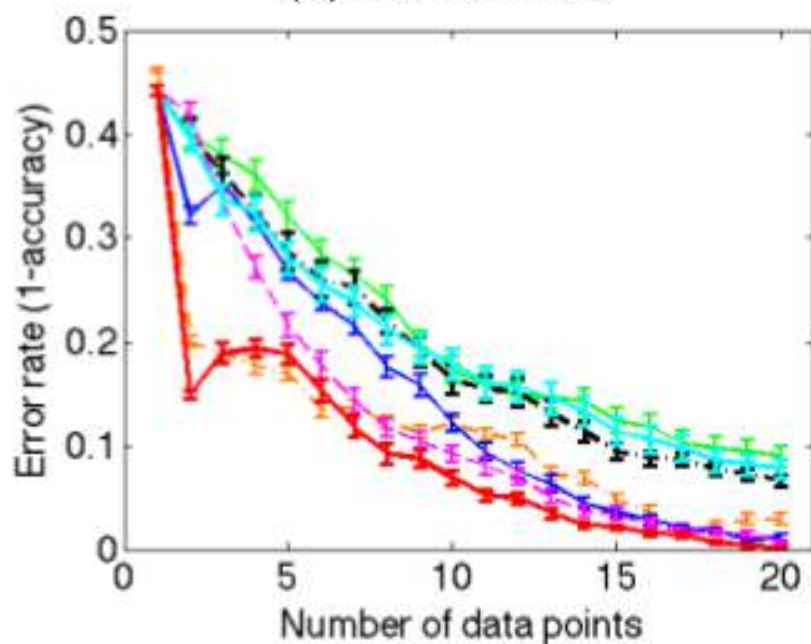
(a) Iris



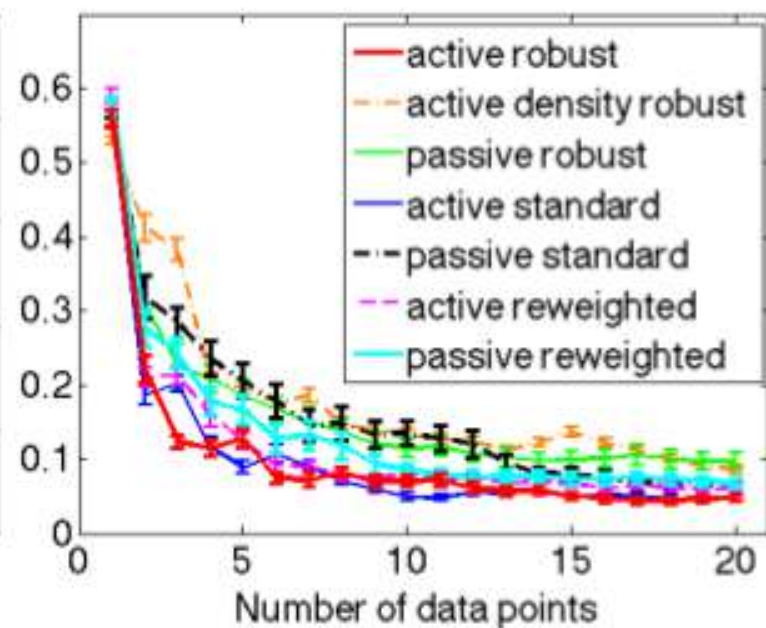
(b) Seed



(c) Banknote

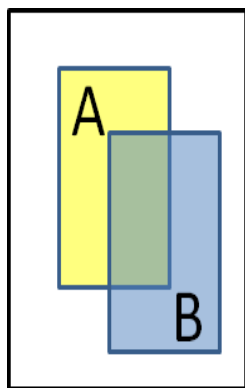
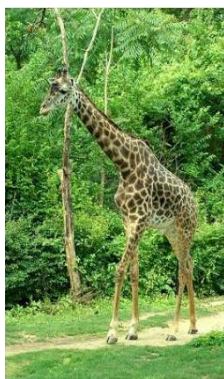


(d) E. coli



# Pessimism Better for Specialized Losses

**Object localization:** performance based on overlap



**Thresholded overlap:**

Incorrect if  $\text{overlap} < 70\%$

Correct if  $\text{overlap} \geq 70\%$

**Methods:** Bayes act under logistic regression (LR) \_  
Hinge loss approximation (SSVM)  
Adversarial object localization (AOL)

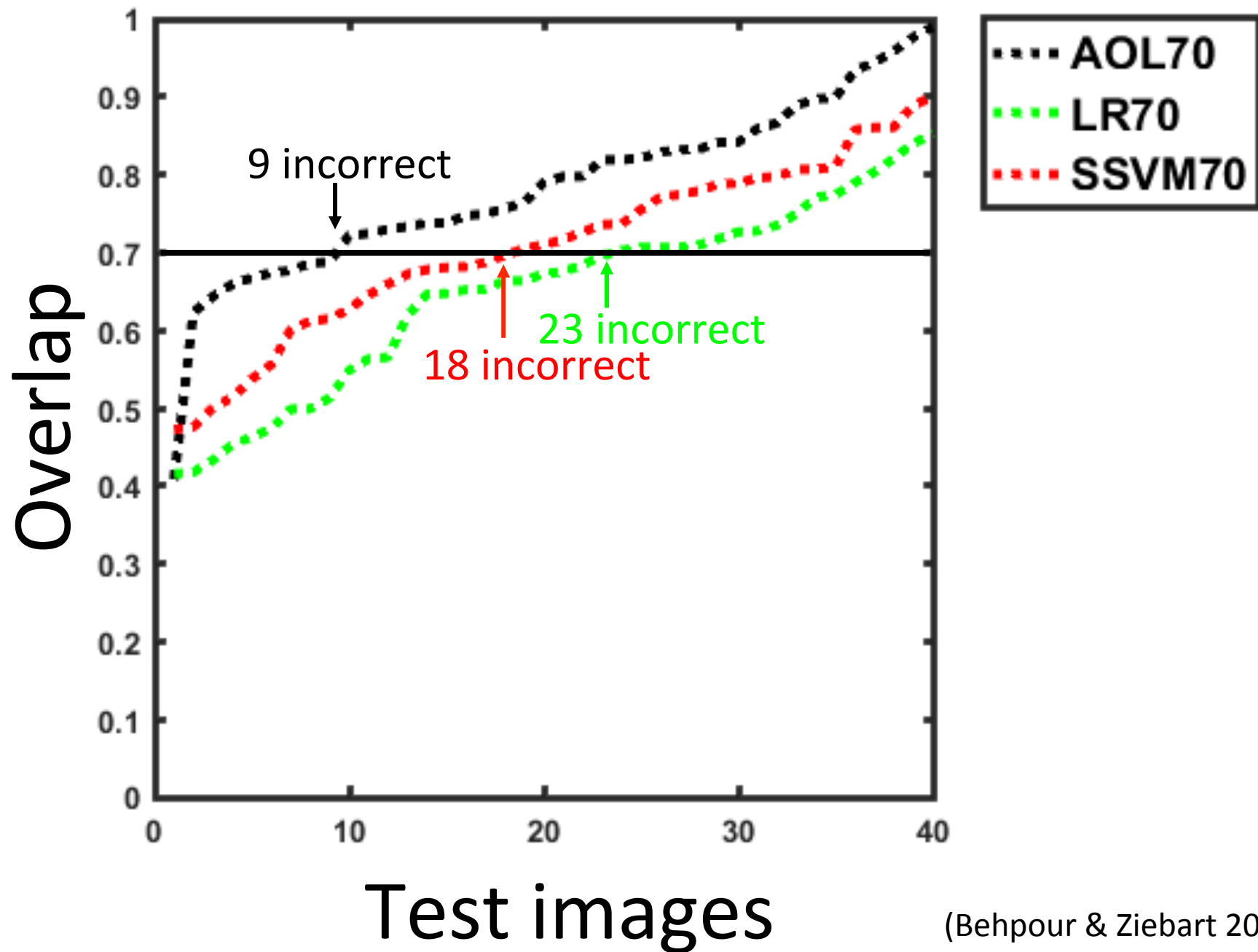
**Experimental setup:** Features from pre-trained deep network (VGGnet); EdgeBox proposals; ImageNet dataset

|           |                                                                                               | Adversary                                                                                       |                                                                                                 |                                                                                                   |                                                                                                   |          |
|-----------|-----------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------|----------|
|           |                                                                                               | $\tilde{y} =$  | $\tilde{y} =$  | $\tilde{y} =$  | $\tilde{y} =$  | ...      |
| Predictor | $\hat{y} =$  | $\ell(\text{tree}, \text{tree}) + \psi(\text{tree})$                                            | $\ell(\text{tree}, \text{giraffe}) + \psi(\text{giraffe})$                                      | $\ell(\text{tree}, \text{giraffe}) + \psi(\text{giraffe})$                                        | $\ell(\text{tree}, \text{giraffe}) + \psi(\text{giraffe})$                                        | ...      |
|           | $\hat{y} =$  | $\ell(\text{giraffe}, \text{tree}) + \psi(\text{tree})$                                         | $\ell(\text{giraffe}, \text{giraffe}) + \psi(\text{giraffe})$                                   | $\ell(\text{giraffe}, \text{giraffe}) + \psi(\text{giraffe})$                                     | $\ell(\text{giraffe}, \text{giraffe}) + \psi(\text{giraffe})$                                     | ...      |
|           | $\hat{y} =$  | $\ell(\text{giraffe}, \text{tree}) + \psi(\text{tree})$                                         | $\ell(\text{giraffe}, \text{giraffe}) + \psi(\text{giraffe})$                                   | $\ell(\text{giraffe}, \text{giraffe}) + \psi(\text{giraffe})$                                     | $\ell(\text{giraffe}, \text{giraffe}) + \psi(\text{giraffe})$                                     | ...      |
|           | $\hat{y} =$  | $\ell(\text{giraffe}, \text{tree}) + \psi(\text{tree})$                                         | $\ell(\text{giraffe}, \text{giraffe}) + \psi(\text{giraffe})$                                   | $\ell(\text{giraffe}, \text{giraffe}) + \psi(\text{giraffe})$                                     | $\ell(\text{giraffe}, \text{giraffe}) + \psi(\text{giraffe})$                                     | ...      |
|           | $\vdots$                                                                                      | $\vdots$                                                                                        | $\vdots$                                                                                        | $\vdots$                                                                                          | $\vdots$                                                                                          | $\ddots$ |

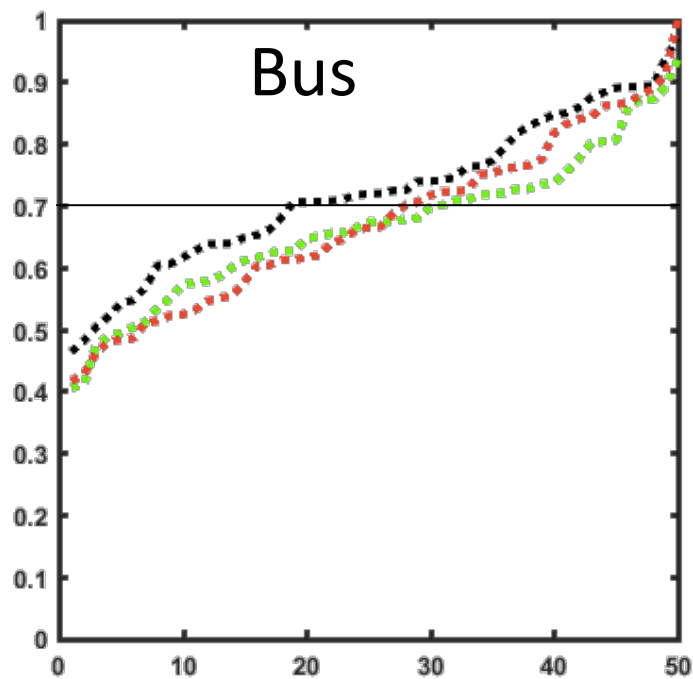
- Prediction game is large (conceptually)
- Efficiently solved using **double oracle method** (McMahan et al. 2003) for constraint generation



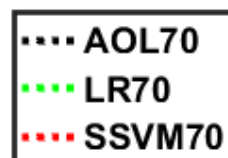
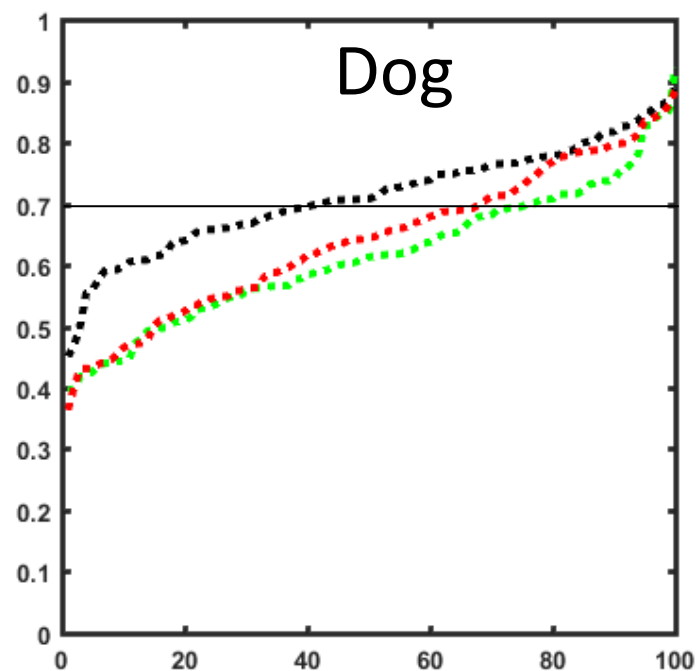
# Cow



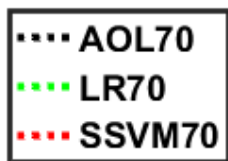
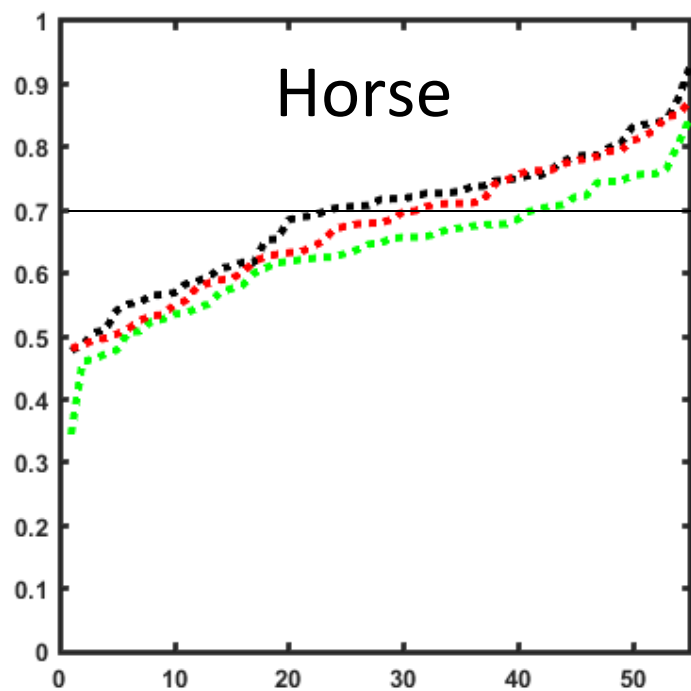
Bus



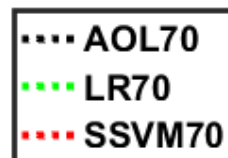
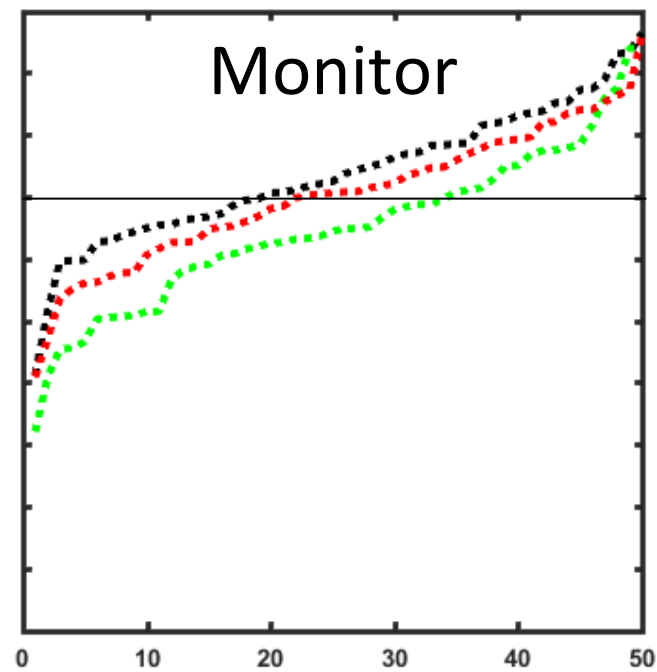
Dog



Horse



Monitor



# Conclusions

**Pessimistic formulation** provides benefits:

- In *theory* (**generalization bounds, consistency**)
- In *practice* (**custom losses, active learning**)

**Questions:**

- Efficiently solve structured predictions games?
- Effective active learning in high dimensions?
- Stronger theory of robust active learning?

## **Related papers:**

Liu, Ziebart. *“Robust Classification Under Sample Selection Bias.”* NIPS 2014.

Liu, Reyzin, Ziebart. *“Shift-Pessimistic Active Learning Using Robust Bias-Aware Prediction.”* AAAI 2015.

Asif, Xing, Behpour, Ziebart. *“Adversarial Cost-Sensitive Classification.”* UAI 2015.

Wang, Xing, Asif, Ziebart. *“Adversarial Prediction Games for Multivariate Losses.”* NIPS 2015.

Li, Asif, Wang, Ziebart. *“Adversarial Sequence Tagging.”* IJCAI 2016.

Chen, Monfort, Liu, Ziebart. *“Robust Covariate Shift Regression.”* AISTATS 2016.

Fathony, Liu, Asif, Ziebart. *“Adversarial Multiclass Classification: A Risk Minimization Perspective.”* NIPS 2016.



NRI Award #1227495  
IIS Award #1526379

