Learning Classification with Auxiliary Probabilistic Information

Quang Nguyen
Hamed Valizadegan
Milos Hauskrecht

Computer Science Department
University of Pittsburgh

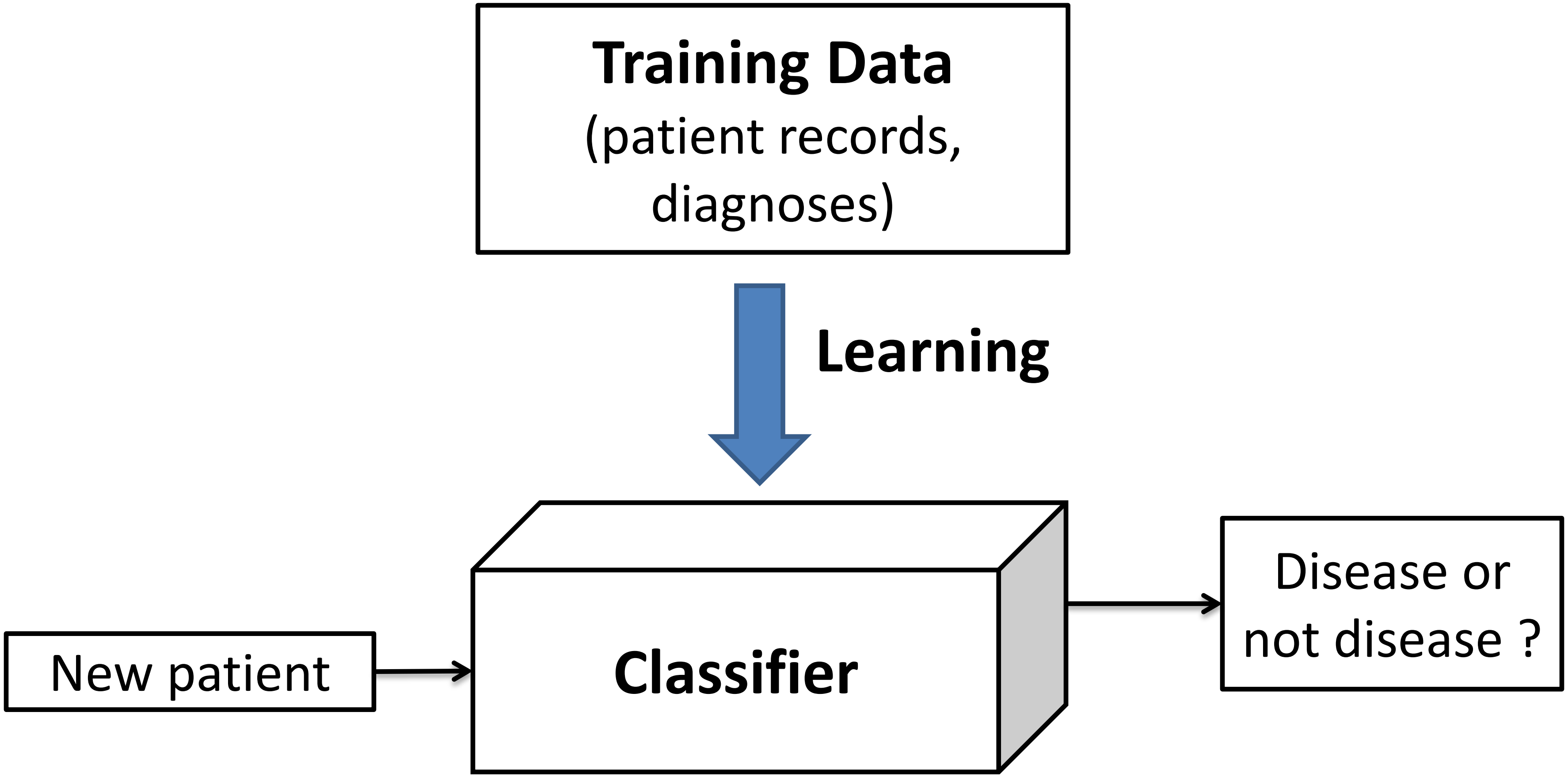
Outline

- Introduction
- Learning with auxiliary information
  - Framework
  - Noise issue
  - Modeling pairwise order constraints
  - Combining class and auxiliary information
- Experimental evaluation
- Conclusion

Outline

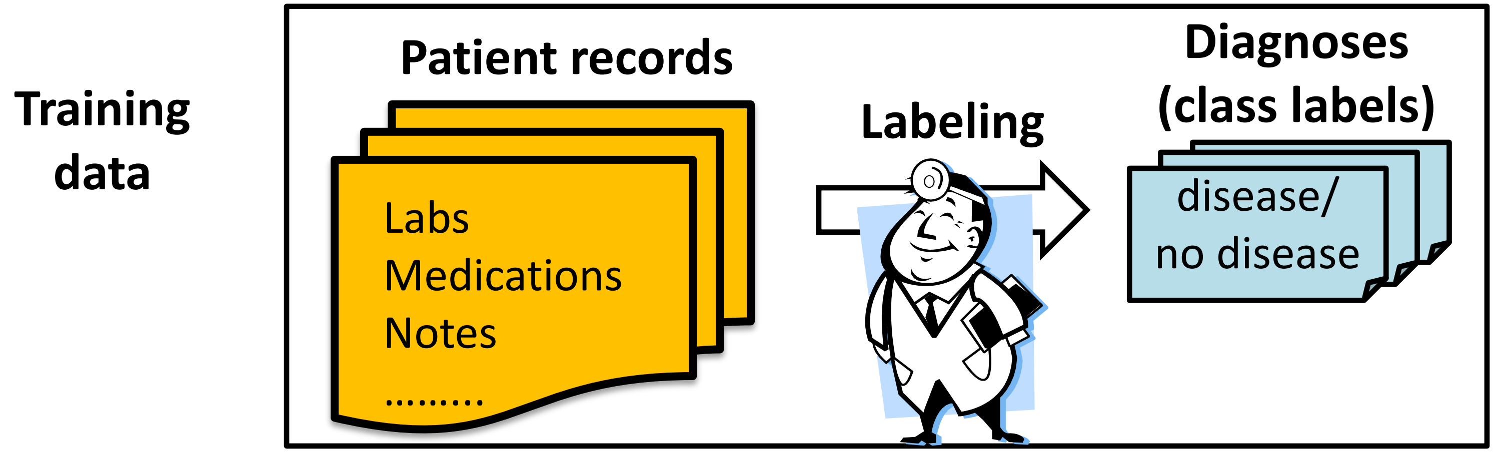
- Introduction
- Learning with auxiliary information
  - Framework
  - Noise issue
  - Modeling pairwise order constraints
  - Combining class and auxiliary information
- Experimental evaluation
- Conclusion

Building a Classification Model



Typically: More training data => Better classifier

Data Labeling and Its Cost



- Labeling requires human experts

$\Rightarrow$ Time consuming and costly

$\Rightarrow$ Small training data

- **How to reduce the number of examples to label ?**
  - Active learning: select only the most critical examples to label
  - **Can we obtain more information from selected examples ?**

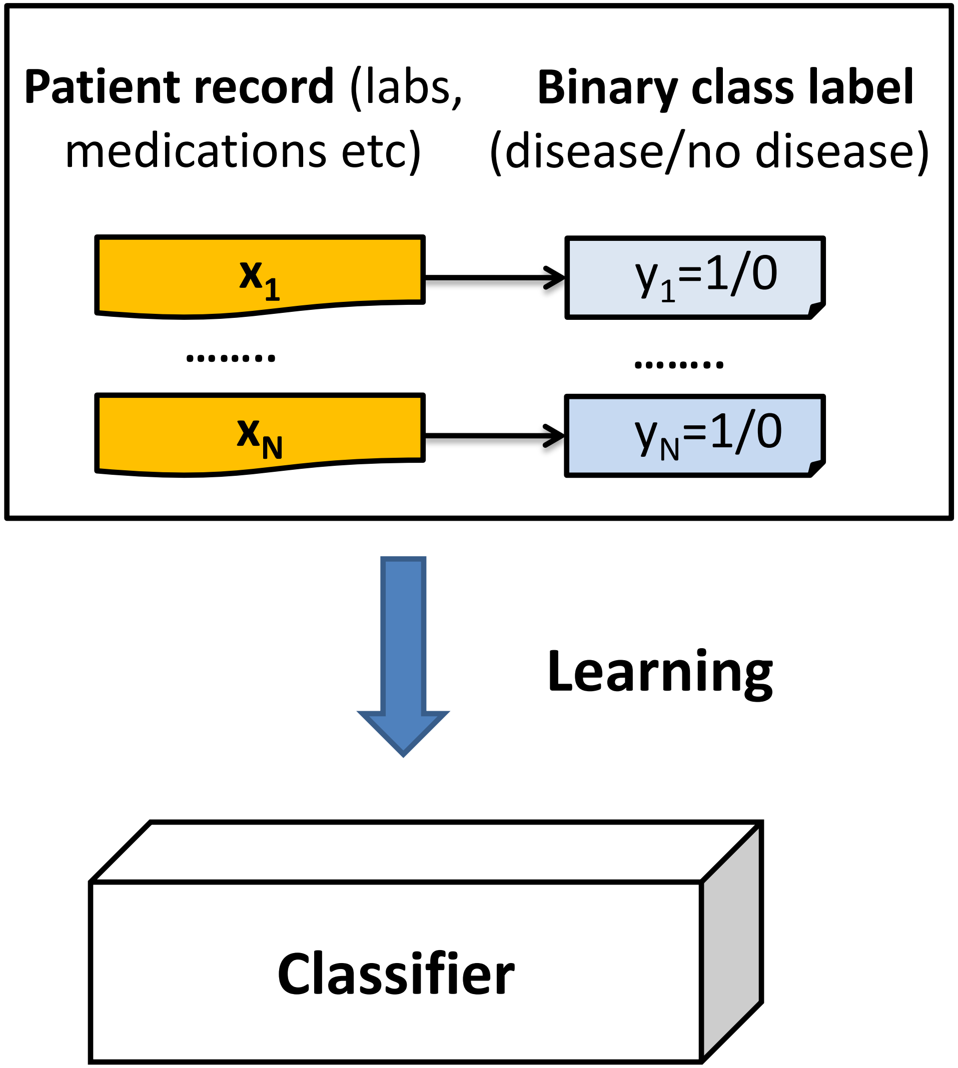
Our Solution

- **Idea**: ask a human expert to provide, in addition to class labels, his/her certainty in the label decision and incorporate this information into the learning process
- Certainty can be represented in terms of
  - Probability: e.g. probability of having disease p = 0.85
  - Qualitative ordinal category: e.g. strong, medium or weak belief in disease; or a discrete score from 0 to 5

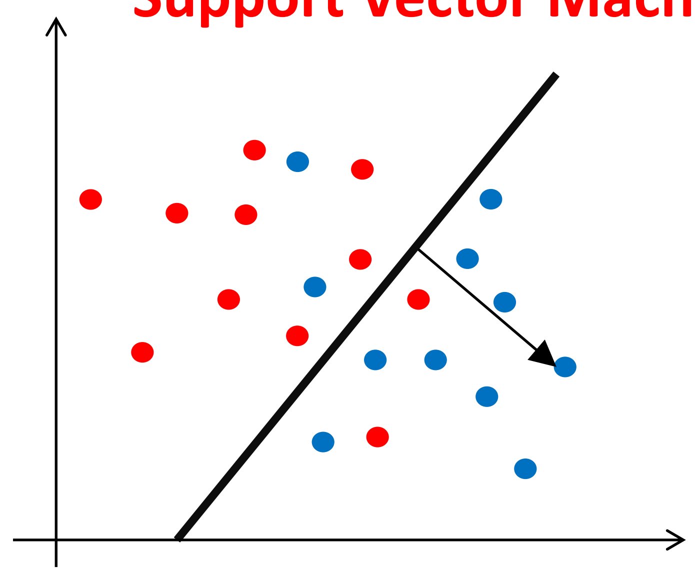
Outline

- Introduction
- Learning with auxiliary information
  - Framework
  - Noise issue
  - Modeling pairwise order constraints
  - Combining class and auxiliary information
- Experimental evaluation
- Conclusion

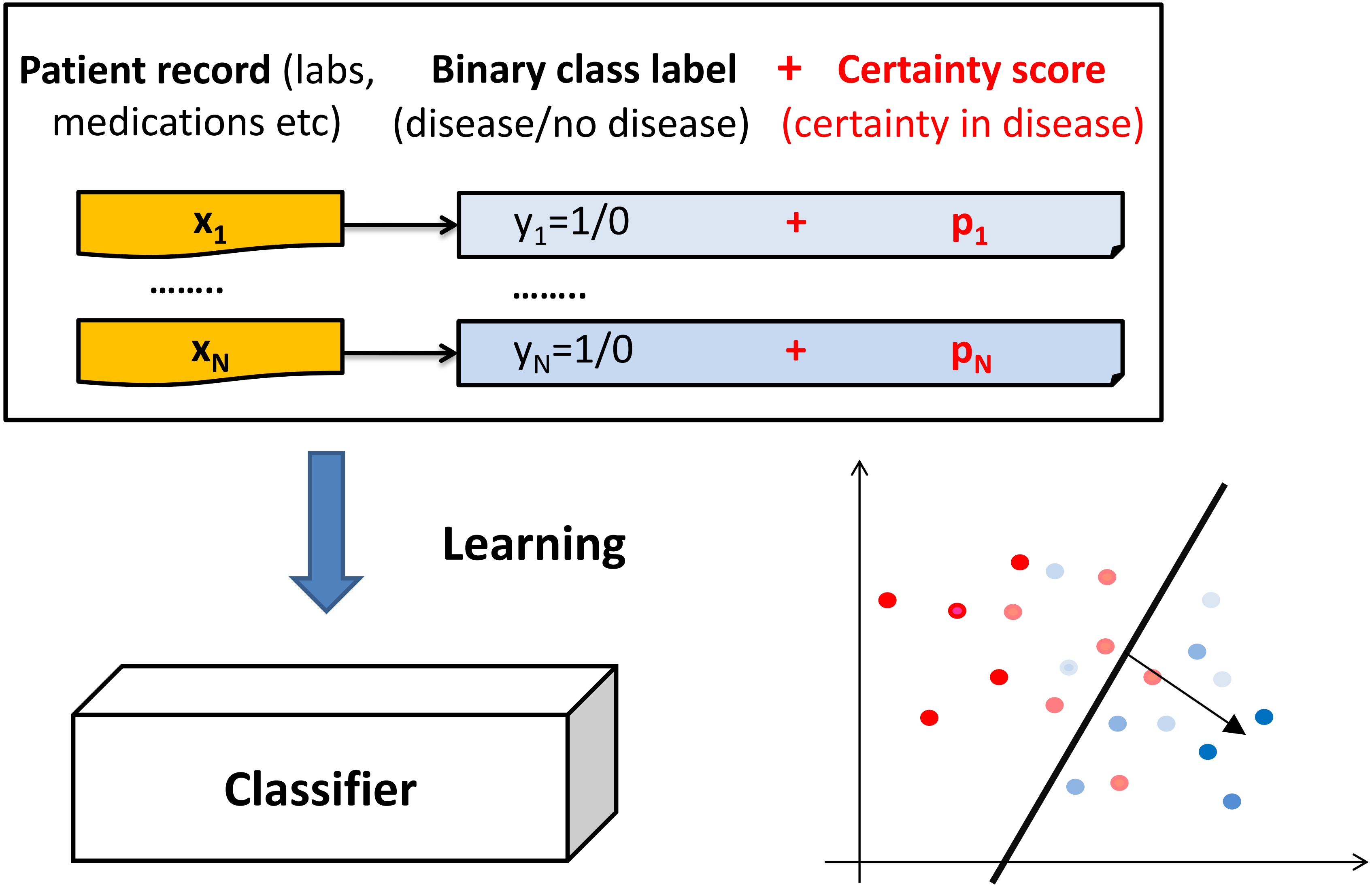
Training with Class Label Information



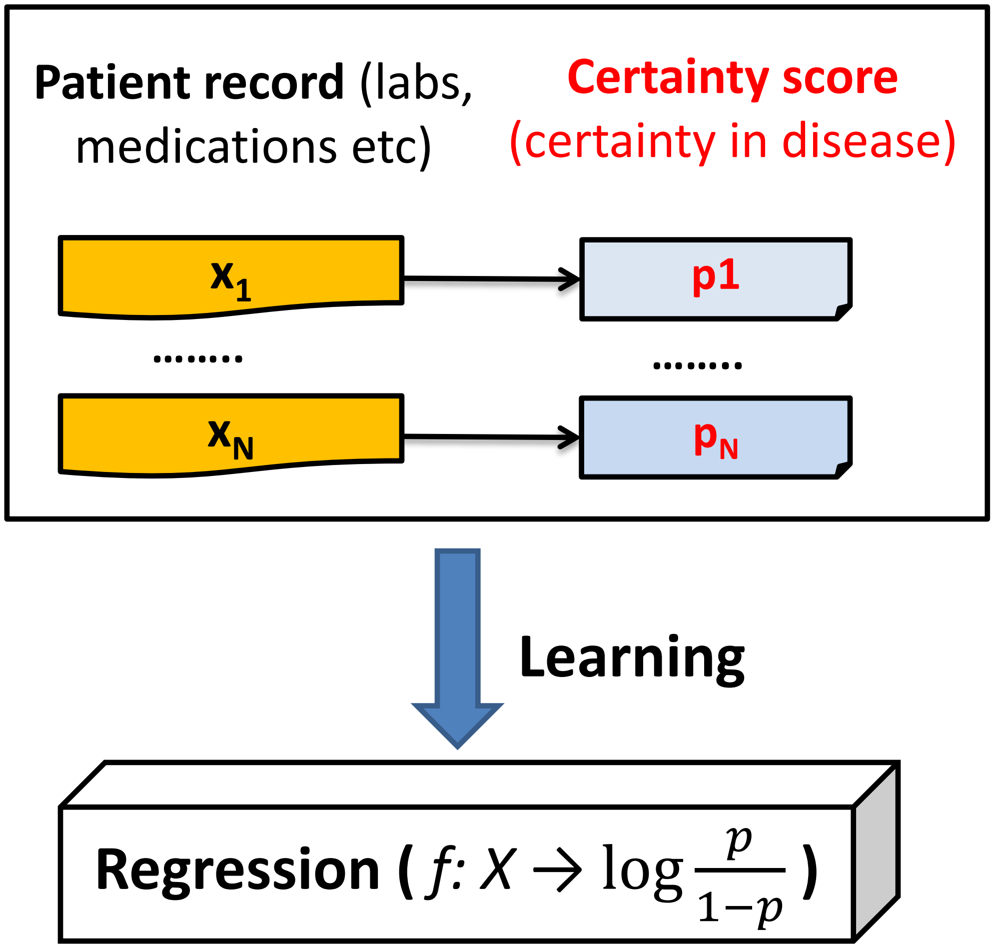
Support Vector Machines

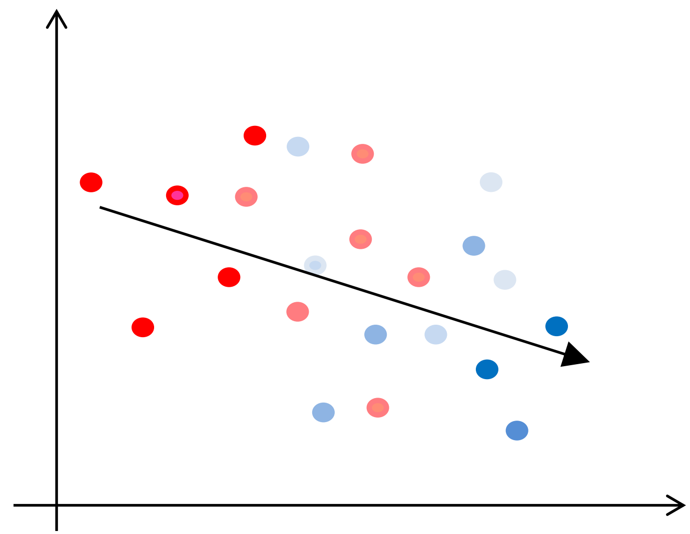


Training with Class Label Information



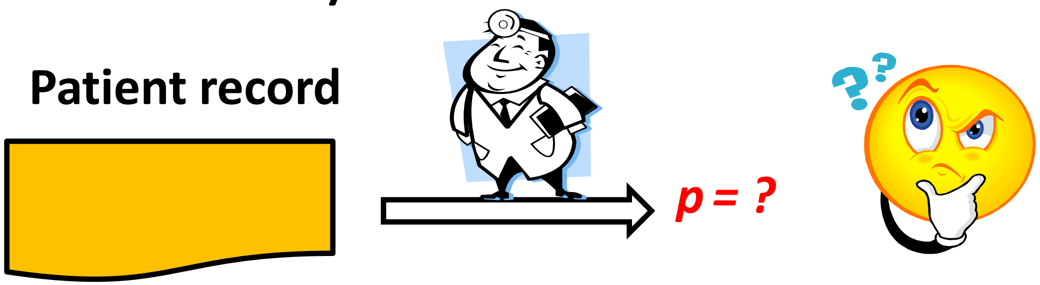
Learning with Auxiliary Information: Regression





Learning with Auxiliary Information: Noise Issue

- Human certainty estimates are often noisy



  - certainty score ***p*** may be inconsistent
- Regression relies on exact values of *p*

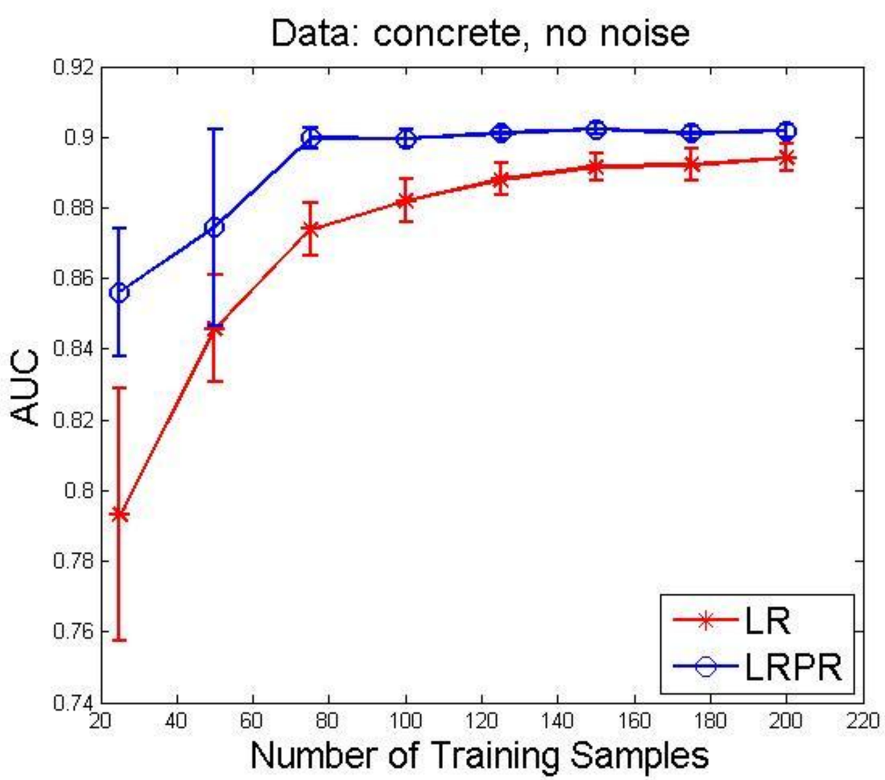
$\Rightarrow$ Sensitive to noise

Learning with Auxiliary Information: Noise Issue

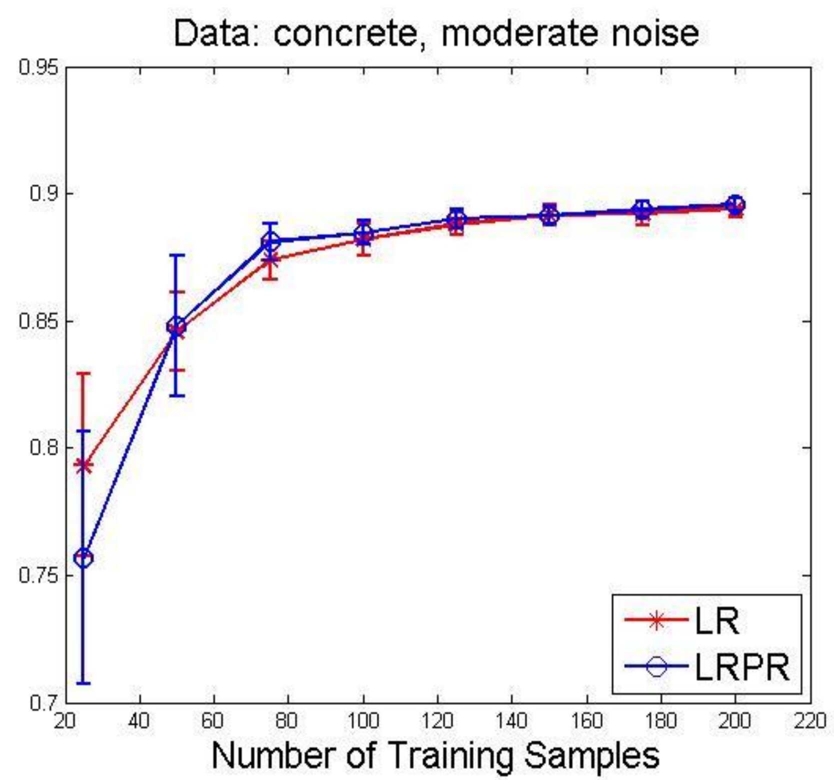
LR: Logistic Regression with binary class labels

LRPR: Logistic Regression with certainty labels

No noise: LRPR clearly outperforms LR



With noise: LRPR is not better than LR

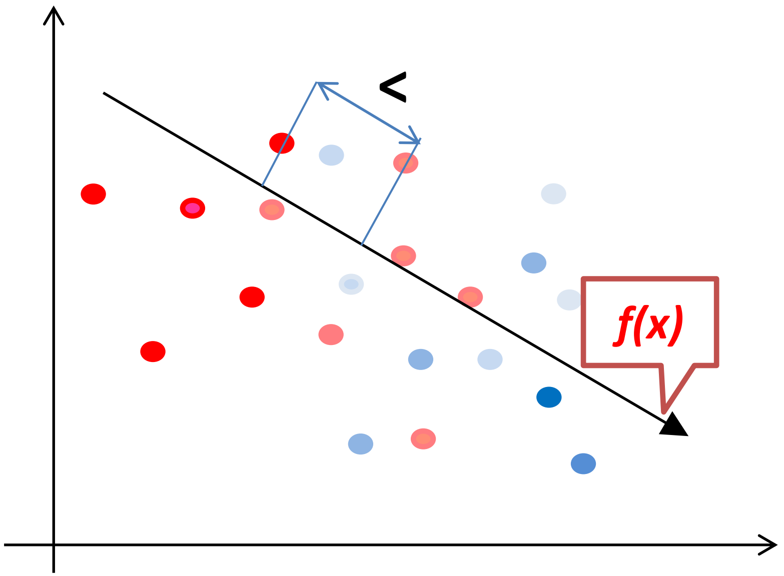


Solution ?

Modeling pairwise orders

- **Observation:** Certainty scores let us order examples
- **Idea:** build a discriminant projection *f(x)* that respects this order
  - Minimize the number of violated pairwise order constraints
- Modeling pairwise orders instead of relying on exact values of p

  => learning less sensitive to noise



Learning with Class and Pairwise Order Constraints

- **Modeling pairwise orders: adapt SVM Rank (Herbrich 2000)**
- **Combining class and certainty information**
  - Optimize:

$$\min_w \frac{1}{2}\mathbf{w}^T\mathbf{w} + \boxed{C \sum_{i,j:p_i>p_j} \xi_{i,j}} + \boxed{B \sum_{i=1}^{N} \eta_i}$$

Penalty for violating class constraints

  - Pairwise order constraints:

$$\forall i,j: p_i > p_j: \quad \mathbf{w}^T(\mathbf{x}_i - \mathbf{x}_j) \geq 1 - \xi_{i,j}$$

$$\forall i \forall j: \quad \xi_{i,j} \geq 0$$

Penalty for violating pairwise orders constraints

  - Class constraints:

$$\forall i: \quad \mathbf{w}^T\mathbf{x}_i y_i + b \geq 1 - \eta_i$$

$$\forall i: \quad \eta_i \geq 0$$

Note: constants B and C regularize the trade-off between class and auxiliary certainty information

Outline

- Introduction
- Learning with auxiliary information
  - Framework
  - Noise issue
  - Modeling pairwise order constraints
  - Combining class and auxiliary information
- Experimental evaluation
- Conclusion

Experimental Setup

- Models
  - Trained on class labels
    - **LR:** standard logistic regression
    - **SVM**: standard linear SVM
  - Trained on certainty labels
    - **LRPR**: Logistic Regression with lasso regularization
  - Trained on both class and certainty labels
    - **SVM-Combo**: SVM with 2 hinge losses for class and pairwise order constraints
- Evaluation
  - Fixed test set
  - Training examples were randomly sampled from train set
  - Repeat training/testing process 30 times
  - Average **AUC** (Area under ROC curve) and 95% confidence interval were recorded

Experimental Setup: UCI Data

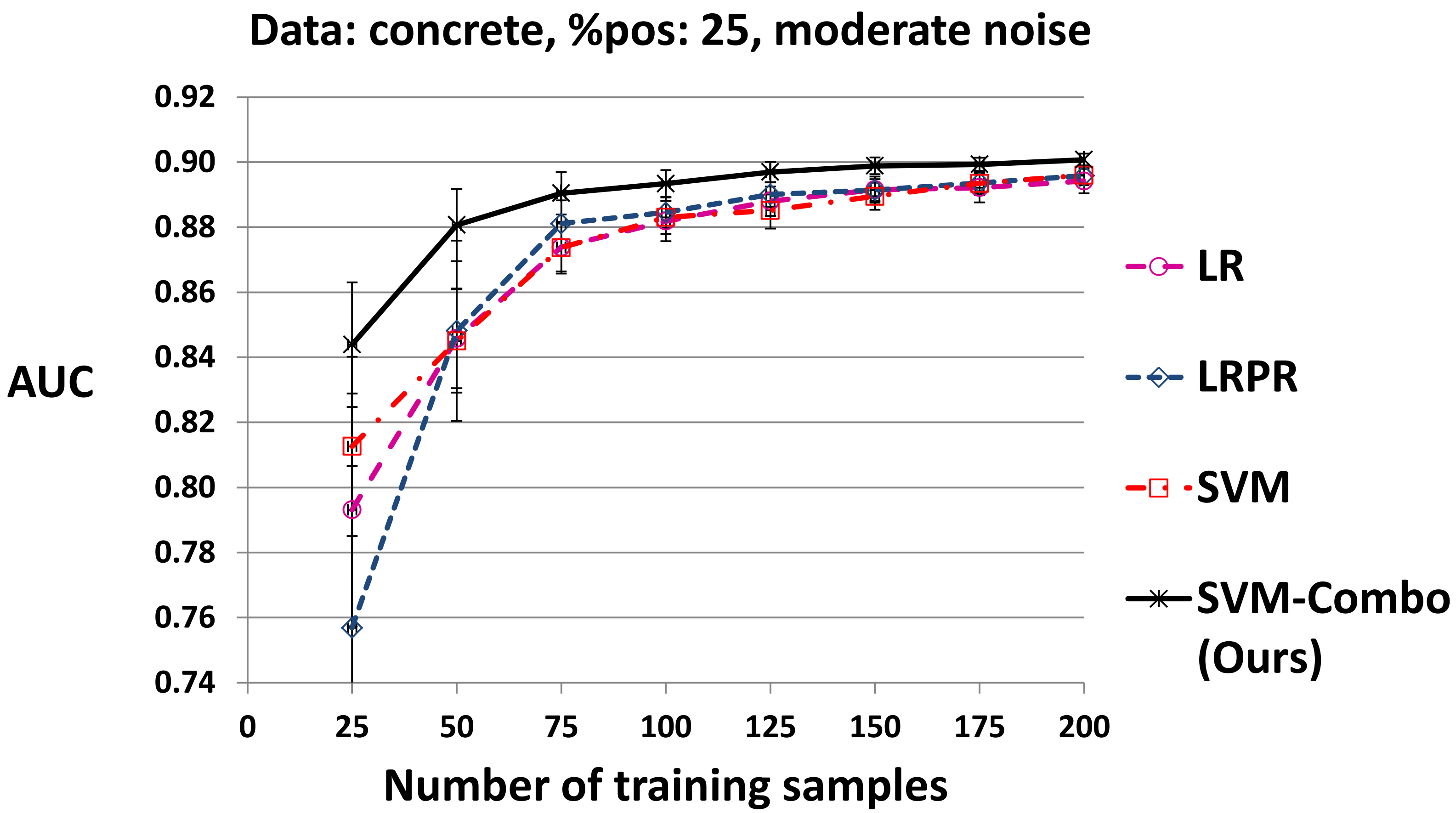
- UCI data sets with continuous outputs
  - Ailerons, Concrete, Kinematics, Puma32
- Generated labels
  - Certainty labels: by normalizing continuous outputs
  - Binary labels: by setting a threshold on certainty labels
- Ratios of positive examples
  - 10%, 25% and 50%

Experimental Setup: UCI Data (Cont'd)

- Noises added to certainty labels
  - 4 different levels: no noise, weak, moderate, strong noises, generated from Gaussian 0.05, 0.1, 0.2 * N(0,1) respectively.
  - Average noise to signal ratios:

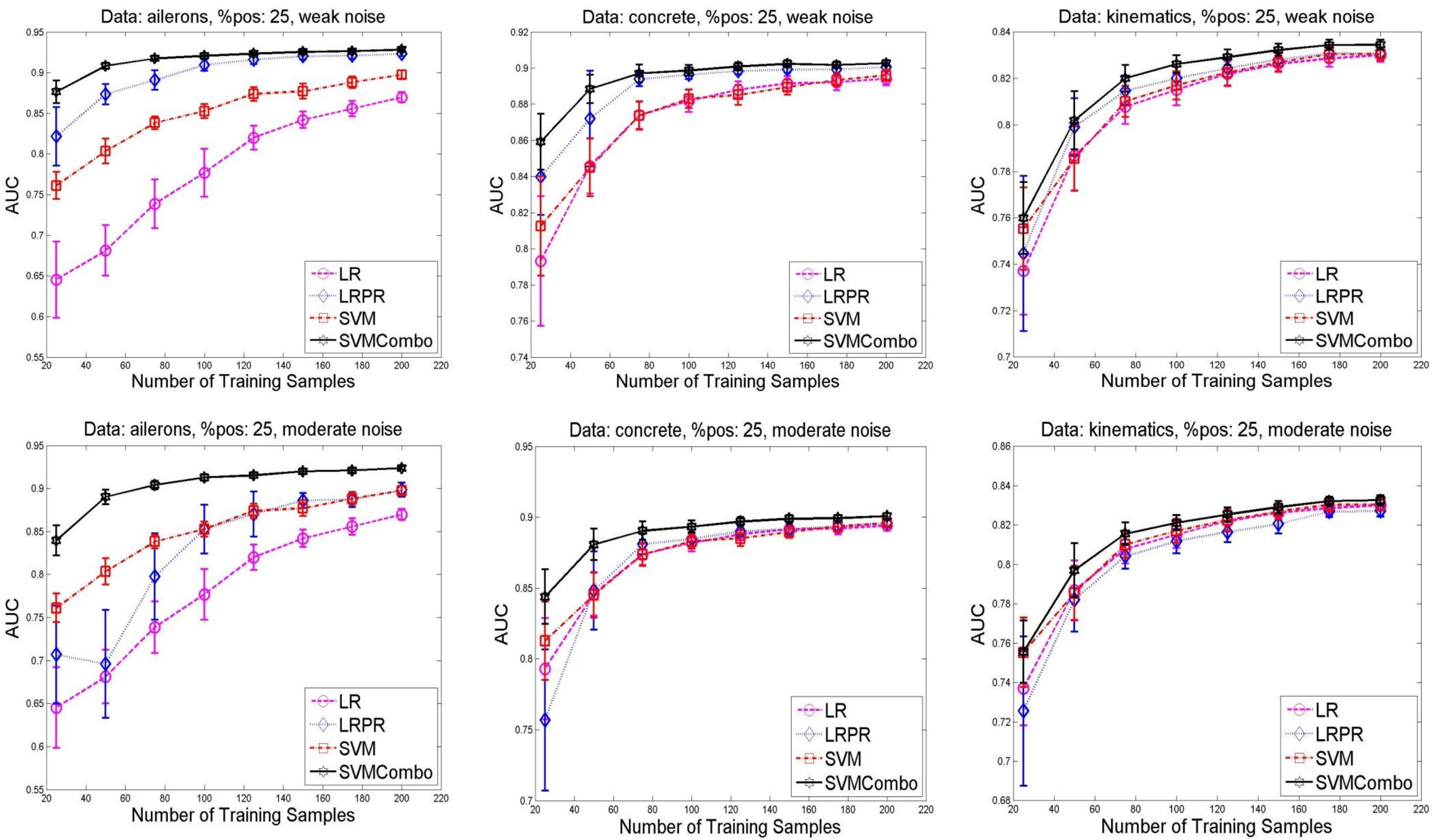
Data Set	Weak Noise	Moderate Noise	Strong Noise
Ailerons	5.2 %	10.3 %	39.8 %
Kinematics	10.6 %	20.8 %	38.9 %
Puma32	10.3 %	20.2 %	39.3 %
Concrete	15.2 %	29.6 %	55.1 %

Experimental Results: UCI Data



- Our method (SVM-Combo) consistently outperforms both regression and standard SVM

Experimental Results: UCI Data (Cont'd)

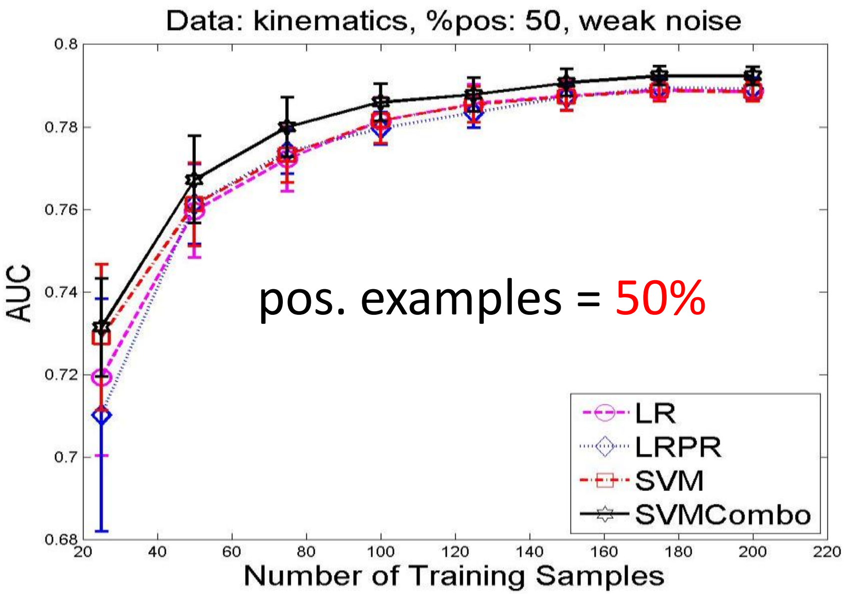


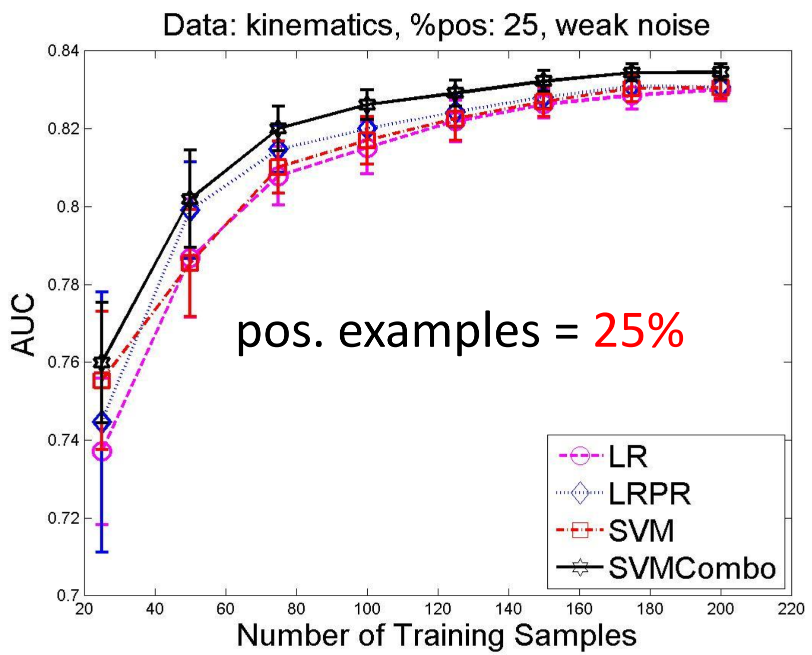
- Our method (SVM-Combo) consistently outperforms both regression and standard SVM

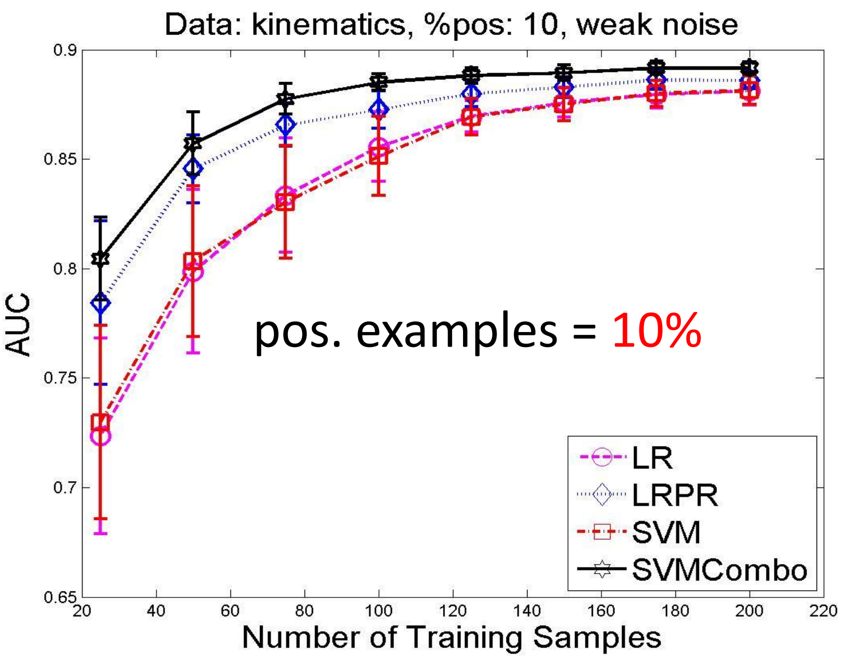
Experiments: Unbalanced Data

- Challenge: in many applications data are often unbalanced (e.g. in medicine positive examples are usually rare)

  Does certainty information help ?
- Auxiliary information shows more benefits with unbalanced data







Experiments: HIT Data

Heparin-induced thrombocytopenia (HIT):

- A life-threatening condition that may develop when patients are treated by heparin

Data:

- Derived from PCP database (Hauskrecht et al. AMIA 2010)
- 199 patient instances labeled by an expert wrt HIT
- 50 features derived from time series of labs, medications and procedures

HIT Data

Labeling:

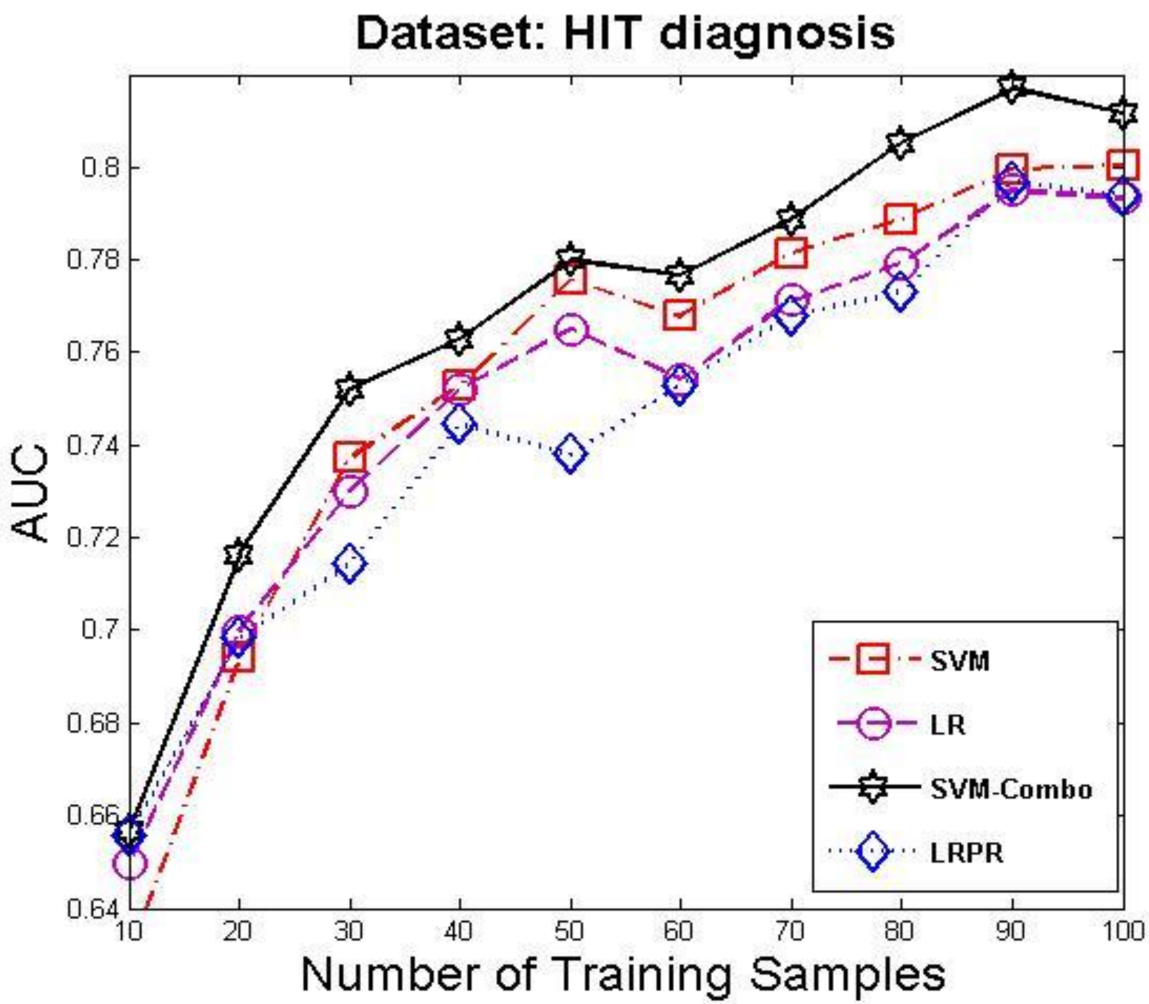
- For each patient case we asked the expert 2 questions
  - Do you agree with raising an alert on HIT or not ? Yes/no
  - How strongly do you agree ? Scale 0 – 4 : strongly-disagree to strongly-agree

Case review:

- We used an in-house EHR graphical interface to collect labels
  - Average time to review a patient: 247 seconds
  - Average time to enter labels: under 10 seconds

$\Rightarrow$ The cost of collecting the auxiliary information is low

Experimental Results: HIT Data



- Our method (SVM-Combo) consistently outperforms both regression and standard SVM

Outline

- Introduction
- Learning with auxiliary information
  - Framework
  - Noise issue
  - Modeling pairwise order constraints
  - Combining class and auxiliary information
- Experimental evaluation
- Conclusion

Conclusion

- Auxiliary certainty information
  - Helps to learn better classification models with smaller numbers of examples
  - Especially useful when data are unbalanced
  - Can be obtained with little additional cost
- Human subjective certainty assessments are noisy
- Proposed method is robust to noise
  - Pairwise orders are more consistent than exact estimates

Thank you for your attention !

- **Contact:** Quang Nguyen, quang@cs.pitt.edu

- **Acknowledgment:** this research was supported by grants from the National Institute of Health
  - 1R01LM010019 (M. Hauskrecht)
  - 1R01GM088224 (M. Hauskrecht)