CS 3750 Machine Learning
Lecture 1

Advanced Machine Learning

Milos Hauskrecht
milos@cs.pitt.edu
5329 Sennott Square, x4-8845

http://www.cs.pitt.edu/~milos/courses/cs3750/

Administration

A seminar course

- **Classes:**
  - **Lectures**
  - **Student (topic-centered) presentations**
- **No homework assignments**
- **Short abstracts for assigned readings due before the class**
- **Course projects**

Administration

Course Projects:

- **2 projects**
  - **Midterm project (assigned)**
  - **Final project (selected areas, student input welcomes)**

- **Grading:**
  - **Projects**
  - **Paper presentations/ discussions**

Study material

Study material:

- **Textbook from CS 2750**
- **Handouts (electronic or hardcopy form)**

- **Books:**
  - Chris Bishop. *Pattern recognition and Machine Learning* Springer, 2006.

Study material

Study material

Other books:

- Koller, Friedman. Probabilistic graphical models.
- Duda, Hart, Stork. *Pattern classification*. 2^{nd} edition. J Wiley and Sons, 2000.
- Friedman, Hastie, Tibshirani. *Elements of statistical learning*. Springer, 2001.
- B. Scholkopf and A. Smola. *Learning with kernels.* MIT Press, 2002.

Tentative topics

- **Review:** **supervised learning, density estimation**
- **Probabilistic models:**
  - **BBNs, MRFs, Monte Carlo inference, variational inference**
- **Low dimensional representation of data**
  - **Component analysis and Latent variable models**
- **Non-parametric models and methods:**
  - **Graph-based kernels for classification and clustering**
- **Extending standard learning frameworks:**
  - **active learning, multi-dimensional learning, transfer learning, learning from multiple annotators**
- **Outlier detection:**
  - **unconditional, conditional**

Supervised learning

Data: $D = \{d_1, d_2, .., d_n\}$ **a set of *n* examples**

$d_i = < \mathbf{x}_i, y_i >$

$\mathbf{x}_i$ is input vector, and y is desired output (given by a teacher)

Objective: learn the mapping $f : X \rightarrow Y$

s.t. $y_i \approx f(x_i)$ for all $i = 1, .., n$

Two types of problems:

- **Regression:** X discrete or continuous $\rightarrow$ Y is **continuous**
- **Classification:** X discrete or continuous $\rightarrow$ Y is **discrete**

Active learning

Data: $D = \{d_1, d_2, .., d_n\}$ **a set of *n* examples**

$d_i = < \mathbf{x}_i, ? >$

$\mathbf{x}_i$ is input vector, and y is missing

Goal: learn the mapping $f : X \rightarrow Y$

s.t. $y_i \approx f(x_i)$ for all $i = 1, .., n$

by asking the user for labels for the different examples in D

Objective: **ask as little examples as possible**

Multi-dimensional classification learning

Data: $D = \{d_1, d_2, .., d_n\}$ **a set of *n* examples**

$d_i = < \mathbf{x}_i, \mathbf{y}_i >$

$\mathbf{x}_i$ is input vector, and $\mathbf{y}_i$ is desired set of outputs

Objective: learn the mapping $f : X \rightarrow Y$

s.t. $\mathbf{y}_i \approx f(\mathbf{x}_i)$ for all $i = 1, .., n$

Caveat: dependences among y components

Transfer learning

Data: $D = \{d_1, d_2, .., d_n\}$ **a set of *n* examples**

$d_i = < \mathbf{x}_i, y_i >$

$\mathbf{x}_i$ is input vector, and y_i is a desired output

Objective: learn the mapping $f : X \rightarrow Y$

s.t.

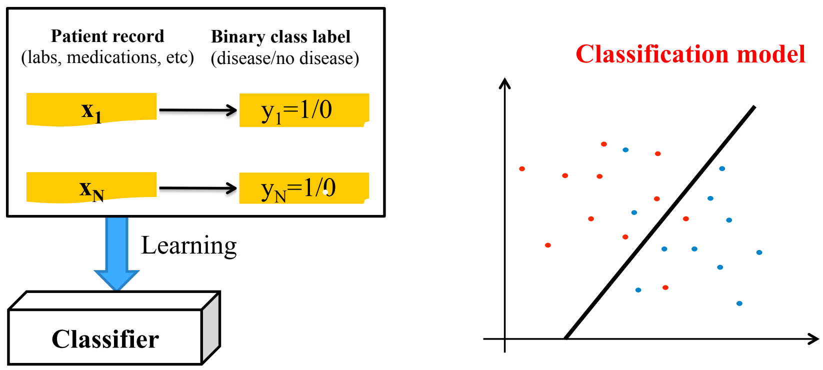
$\mathbf{y}_i \approx f(\mathbf{x}_i)$ for all $i = 1, .., n$

Caveat: n is small and the dimensionality of $\mathbf{x}$ is high

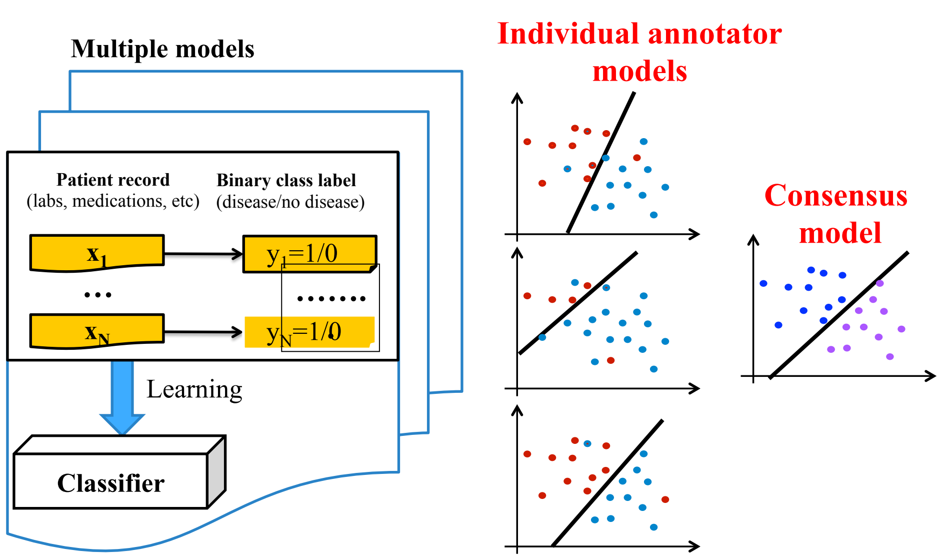
Assumption: we have $D' = \{d_1', d_2', .., d_m'\}$

where $d_i' = < \mathbf{x}_i, z_i >$ and z and y are related

Learning single classification model



Learning with multiple annotators



Anomaly Detection

- **Traditional (unconditional) anomaly detection:**
  - Data D={**x**}
  - **Goal:** find anomalous entries in the data
    - find data entries **x** in low density regions
    - p(**x**) is small relative to other entries
- **Conditional anomaly detection:**
  - Data D={(**x**,**y**)}
  - **Goal:** find entries with anomalous response **y** for **x**
    - find (**x**,**y**) entries for which p(**y**|**x**) is small relative to alternative responses **y** for **x**