Lecture II

Sam Reid
Advisor Greg Grudic

University of Colorado at Boulder
Department of Computer Science

January 30, 2008

Questions from Last Week?

- Questions about last week's homework?
- Questons about Matlab?
- Questions about last week's lecture?

Last Time

- Difference between a classifier and a classification algorithm?
- Difference between parameters and hyperparameters? Examples?
- Why can't the training set be used to perform model assessment?
- Relationship between model selection and model assessment?
- How to implement k-fold cross-validation?
- What is a loss function?
- How to compute mean-squared error?
- What is stratification? Why is it used?

Supervised Classification

- Obtain dataset $\mathcal{D} = \{\mathbf{z}_i\}_{i=1}^N$, $\mathbf{z}_i = (\mathbf{x}_i, y_i)$, $\mathbf{x}_i \in \mathbb{R}^d$, $y_i \in \Omega = \{\omega_1, ..., \omega_c\}$
- Search for $\hat{y}(\mathbf{x}|\theta) : \mathbb{R}^d \Rightarrow \Omega$ to minimize future risk:

$$R(F) = E_{q(\mathbf{z})} L(y, F(\mathbf{x}))$$

- $L(y, \hat{y})$ is the loss function.

Loss Functions

- Squared Error (regression):

$$L = (\hat{y} - y)^2$$

- Zero-One Loss (classification):

$$L = 1 - \delta(\hat{y} - y)$$

- Can also be input-point dependent: $L(y, \hat{y}, \mathbf{x})$

Risk

- Expectation value of the loss function:

$$R(F) = E_{q(\mathbf{z})} L(y, F(\mathbf{x}))$$

- Expectation value:
  - Discrete: $E(X) = \sum_x xP(x)$
  - Continuous: $E(X) = \int xp(x)dx$
- To estimate the expectation value over a finite sample size, assign a probability mass of 1 to each hold-out point.

$$R(F) = \frac{1}{N} \sum_{i=1}^{N} L(y_i, F(\mathbf{x}_i))$$

Bayes' Optimal Error

- a.k.a Bayes' Risk
- All problems have a minimum achievable risk
- For some problems this is 0% error.
- But others have inherent noise: e.g. data points A and B may have the same input features, but different classes.
- The goal is not to get 100%.
- Tumer and Ghosh 1996 "Estimating the Bayes Error Rate through Classifier Combining")
- Problem unsolved in general

Netflix Prize

Weka

New Homework

Questions?

Preprocessing for Machine Learning

- Step 1: Decide on train/test partition of dataset
  - Train partition includes validation (i.e. model selection) data.
  - Stratify: distribution of classes should be same in each partition.
- Step 2: Compute mean and variance of each numeric attribute based on training partition.
- Step 3: Standardize the entire dataset (train + test) by subtracting the train-mean and dividing by the train-standard deviation

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2}$$