[10-701] Introduction to Machine Learning(PhD)

Brandon Jin

Spring 2018

Contents

1 Probability Models

1.1 MLE, Estimators, Guarantees

1.2 MAP, Bayesian Estimation

2 Model-free Methods, Decision Theory

3 Regression

3.1 Linear Regression

3.2 Regularized, Polynomial, Logistic Regression

4 Classification

4.1 Naive Bayes, Generative vs Discriminative

Conditional Independence

X is conditionally independent of Y given Z:
P(X—Y, Z) = P(X—Z). Also, P(X, Y—Z) = P(X—Z)P(Y—Z).
For example, P(Thunder—Rain, Lightning) = P(Thunder—Lightning). Thunder is not independent of rain. However, thunder is **conditionally independent** of rain given lightning. If you see lightning, thunder and rain are independent. Seeing rain does not given you information about thunder in that case.

Naive Assumption: features are independent given class
$P(X_1, X_2|Y) = P(X_1|X_2, Y)P(X_2|Y) = P(X_1|Y)P(X_2|Y)$. More generally, $P(X_1...X_d|Y) = \Pi_{i=1}^{d} P(X_i|Y)$.
Note: if conditional independence assumption hold, Naive Bayes is the optimal classifier. But worse otherwise.

4.1.1 Generative vs Discriminative Model

Generative(Model-BASED) Approach

Discriminative(Model-FREE) Approach

4.2 Support Vector Machines (SVM)

4.3 Boosting, Surrogate Losses

5 Decision Tree

6 Neural Networks and Deep Learning

For each neuron,
Weight tells you what pattern this neuron in the second layer is picking up on **Bias** tells you how high the weighted sum needs to be before the neuron starts getting meaningfully active

7 Non-parametric Models

7.1 K-Nearest Neighbors, Kernel Density Estimation

7.2 SVM, Linear Regression: primal + dual, Kernel Trick

8 Generalization, Model Selection

8.1 True Risk vs Empirical Risk

8.2 Estimating True Risk

8.3 Improving Empirical Risk Minimization

8.4 Model Selection by Estimating True Risk

8.5 Analyzing Generalization Error Via True Risk