Quiz 4

EEE 4774 & 6777 Data Analytics

1 Gaussian Generative Model, QDA, LDA, Gaussian Naive Bayes

Consider the binary classification problem with d-dimensional training data points $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ and labels $\{y_1, \ldots, y_N\}$. Starting with the Gaussian generative model for the two classes, $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ (likelihood for class 1), $p(C_1)$ (prior for class 1), $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ (likelihood for class 2), $p(C_2)$ (prior for class 2), answer the following questions.

a) Derive the Quadratic Discriminant Analysis (QDA) classifier, which compares $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} + \boldsymbol{x}^T\boldsymbol{b}$ with a threshold to make the classification decision (i.e., derive the expressions for $\boldsymbol{A}$, $\boldsymbol{b}$, and the threshold). [20 pts]

b) Show the assumption to derive the Linear Discriminant Analysis (LDA) from QDA. LDA chooses class 1 if $\boldsymbol{x}^T\boldsymbol{w} > h$ and class 2 otherwise. Show the expressions for $\boldsymbol{w}$ and the threshold h. [20 pts]

c) Through another simplifying assumption on the QDA we can obtain the Gaussian Naive Bayes classifier. Explain what that assumption is and how it yields the Gaussian Naive Bayes classifier. [20 pts]

Extra Credit: Show the final simplified (***hint:*** univariate) decision function of the Gaussian Naive Bayes classifier. [15 pts]

2 SVM

Explain the SVM algorithm considering its objective, loss function, kernel trick, training procedure, and testing procedure. [20 pts]

3 Ensemble Methods

Explain the differences between the boosting and bagging methods. [20 pts]