

P9120 - Homework # 3

Assigned: October 24, 2019

Due: November 7, 2019 (in class)

Maximum points that you can score in this Homework is 20.

1. Suppose $X \in \mathbb{R}^p$ and $Y \in \{-1, 1\}$. For any real-valued function f on \mathbb{R}^p , let $L(Y, f(X))$ denote the loss function for measuring errors between Y and $f(X)$. Let $f^* = \arg \min_f EL(Y, f(X))$, where the expectation is taken over the joint distribution of X and Y . Show that

(a) (Logistic Regression) If $L(y, f(\mathbf{x})) = \log[1 + \exp(-yf(\mathbf{x}))]$, then $f^*(\mathbf{x}) = \log \frac{Pr(Y=1|X=\mathbf{x})}{Pr(Y=-1|X=\mathbf{x})}$.

(b) (SVM) If $L(y, f(\mathbf{x})) = [1 - yf(\mathbf{x})]_+$, then $f^*(\mathbf{x}) = \text{sign}[Pr(Y = 1|X = \mathbf{x}) - \frac{1}{2}]$.

(c) (Regression) If $L(y, f(\mathbf{x})) = [y - f(\mathbf{x})]^2$, then $f^*(\mathbf{x}) = 2Pr(Y = 1|X = \mathbf{x}) - 1$.

(d) (AdaBoost) If $L(y, f(\mathbf{x})) = \exp[-yf(\mathbf{x})]$, then $f^*(\mathbf{x}) = \frac{1}{2} \log \frac{Pr(Y=1|X=\mathbf{x})}{Pr(Y=-1|X=\mathbf{x})}$.

2. Get the “Ripley dataset” (synth.tr) from the website <http://www.stats.ox.ac.uk/pub/PRNN/>.

The dataset contains two predictors and a binary outcome.

- (a) Construct a linear support vector classifier.
- (b) Construct a support vector classifier with Radial kernel.
- (c) Construct a classifier using AdaBoost algorithm (with 50 boosting iterations) with decision stumps as weak learners.

Select the tuning parameter involved in SVM models appropriately. For each method, compute the test error and its standard error on the test set (synth.te). Provide a simple graphical visualization of the produced classification models (i.e. something similar to Figure 2.2 in the textbook [ESL]) and discuss your results.