P9120 - Homework # 3

Assigned: October 27, 2022 Due at 1pm on November 10, 2022

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. Ex 10.4, parts (b), (c), (d) of [ESL] (page 385). You can use existing AdaBoost package or write your own program to implement AdaBoost.

3. The "spam" data" (https://web.stanford.edu/ hastie/ElemStatLearn/data) has been divided into a training set and a test set. Fit a neural network to the training set, and calculate its classification error on the test set. Compare your results to the classification tree results presented in Section 9.2.5 of [ESL] on both the classification performance and interpretability of the final model.