CS57800 Statistical Machine Learning
Homework 2

Ting Zhang
School of Industrial Engineering
zhan1013@purdue.edu

September 23, 2015

1 Foundations

1. (1) Boolean function
   $f(x_1, x_2, x_3, x_4) = \neg[(x_1 \wedge \neg x_2 \wedge \neg x_3 \wedge \neg x_4) \vee (\neg x_1 \wedge x_2 \wedge \neg x_3 \wedge \neg x_4) \vee (\neg x_1 \wedge \neg x_2 \wedge x_3 \wedge \neg x_4) \vee (\neg x_1 \wedge \neg x_2 \wedge \neg x_3 \wedge x_4) \vee (\neg x_1 \wedge \neg x_2 \wedge \neg x_3 \wedge \neg x_4)]$
   (2) Linear function $f(x_1, x_2, x_3, x_4) = 1 if x_1 + x_2 + x_3 + x_4 \geq 2$

2. $size(CON_B) = 2^n$

3. Since $\beta_n = \beta_o + y_i u_i$, then

$$\begin{aligned}\|\beta_n - \beta^*\|^2 &= \|\beta_o + y_i u_i - \beta^*\|^2 \\ &= (\beta_o - \beta^*)^2 + 2 y_i u_i (\beta_o - \beta^*) + (y_i u_i)^2 \\ &= (\beta_o - \beta^*)^2 + 2(\beta_o y_i u_i - \beta^* y_i u_i) + (y_i u_i)^2\end{aligned}$$

   Since $u_i = {x_i}^* / \|{x_i}^*\|$, then $u_i = \pm 1$. Also, because u_i is a misclassified example, then $y_i \cdot (\beta_o u_i) = -1$. Because β^* is the final separating parameter vector, we will have $y_i \cdot (\beta^* u_i) = 1$. Substitute these values into the above equation, we will have:

$$\begin{aligned}\|\beta_n - \beta^*\|^2 &= \|\beta_o - \beta^*\|^2 + 2(-1-1) + 1 \\ &= \|\beta_o - \beta^*\|^2 - 3 \\ &\leq \|\beta_o - \beta^*\|^2 - 1\end{aligned}$$

4.

5. (1) Both classifiers will converge since the data is linearly separable. (2) The training error of the first classifier would be 50%. And the training error for the second classifier would be 0% since the data is linearly separable.

6. Proof.

$$\begin{aligned}
\|\sum_{i\in N} y_i x_i\| &= \|\sum_{i\in N}(w_{i+1} - w_i)\| \\
&= \|(w_{i+1} - w_i) + (w_i - w_{i-1}) + (w_{i-1} - w_{i-2}) + ... + (w_1 - w_0)\| \\
&= \|w_{i+1}\| \\
&= \sqrt{(\|w_{i+1}\|^2 - \|w_i\|^2) + (\|w_i\|^2 - \|w_{i-1}\|^2) + ... + (w_1^2 - w_0^2)} \\
&= \sqrt{\sum_{i\in N}(\|w_{i+1}\|^2 - \|w_i\|^2)} \\
&= \sqrt{\sum_{i\in N}(\|w_i + y_i x_i\|^2 - \|w_i\|^2)} \\
&= \sqrt{\sum_{i\in N}(\|w_i\|^2 + 2y_i w_i x_i + \|y_i x_i\|^2 - \|w_i\|^2)} \\
&= \sqrt{\sum_{i\in N}(2y_i w_i x_i + \|x_i\|^2)}
\end{aligned}$$

Since these are the examples when Peceptron makes a mistake, therefore, $y_i w_i x_i \leq 0$. So,

$$\|\sum_{i\in N} y_i x_i\| \leq \sqrt{\sum_{i\in N} \|x_i\|^2}$$

2 Programming Report

...