Stats 441 - Homework 4

Junqiao Lin (20609800)

December 3, 2018

PROBLEM 1.

a)
Note by definition, if $\hat{y} = 0$, we have

$$P(y=1)\Pi j = 1^n P(x_j|y=1) \le P(y=0)\Pi_{j=1}^n P(X_j|y=0)$$

$$\log \frac{P(y=1)}{P(y=0)} + \Sigma_{j=1}^n \log(p(x_j|y=1)) - \log(p(x_j|y=0)) \le 0$$

Note since $x_j = 1$ we have $p(x_j|y=1) = P_{j1}$, $p(x_j|y=0) = P_{j0}$, $x_j = 0$ we have $p(x_j|y=1) = 1 - P_{j1}$, $p(x_j|y=0) = 1 - P_{j0}$, thus we have:

$$\log \frac{p}{1-p} + \Sigma_{j=1}^n ((\log(p_{j1}) + \log(1-p_{j0}) - \log(p_{j0}) - \log(1-p_{j0}))x_j + (\log(1-p_{j1}) - \log(1-p_{j0}))) \le 0$$

$$(\Sigma_{j=1}^n ((\log(p_{j1}) + \log(1-p_{j0}) - \log(p_{j0}) - \log(1-p_{j0}))x_j) + (\Sigma_{j=1}^n (\log(1-p_{j1}) - \log(1-p_{j0}))) + \log \frac{p}{1-p}) \le 0$$

Thus we have

$$\begin{aligned} w_j &= (\log(p_{j1}) + \log(1-p_{j0}) - \log(p_{j0}) - \log(1-p_{j0})) \\ &= \log(\frac{p_{j1}(1-p_{j0})}{(1-p_{j1})(p_{j0})}) \\ b &= (\Sigma_{j=1}^n (\log(1-p_{j1}) - \log(1-p_{j0}))) + \log \frac{p}{1-p} \\ &= \log(\frac{p}{1-p} \Pi_{i=1}^n \frac{1-p_{j1}}{1-p_{j0}}) \end{aligned}$$

b)
Note by definition, if $\hat{y} = 0$, we have

$$P(y=1)\Pi j = 1^n P(x_j|y=1) \leq P(y=0)\Pi_{j=1}^n P(X_j|y=0)$$

$$\log\frac{P(y=1)}{P(y=0)} + \Sigma_{j=1}^n (P(x_j|y=1) - \log(p(x_j|y=0))) \leq 0$$

$$\log\frac{p}{1-p} + \frac{1}{2}\Sigma_{j=1}^n (\frac{(x_j-\mu_{j1})^2 - (x_j-\mu_{j0})^2}{\sigma_j^2}) \leq 0$$

$$\log\frac{p}{1-p} + \frac{1}{2}\Sigma_{j=1}^n (\frac{x_j^2 - 2x_j\mu_{j1} + \mu_{j1}^2 - x_j^2 + 2x_j\mu_{j0} - \mu_{j0}^2}{\sigma_j^2}) \leq 0$$

$$\log\frac{p}{1-p} + \frac{1}{2}\Sigma_{j=1}^n (\frac{-2x_j\mu_{j1} + \mu_{j1}^2 + 2x_j\mu_{j0} - \mu_{j0}^2}{\sigma_j^2}) \leq 0$$

$$\log\frac{p}{1-p} + \frac{1}{2}\Sigma_{j=1}^n (\frac{(2\mu_{j0} - 2\mu_{j1})x_j}{\sigma_j^2} + \frac{\mu_{j1}^2 - \mu_{j0}^2}{\sigma_j^2}) \leq 0$$

$$\Sigma_{j=1}^n (\frac{(\mu_{j0} - \mu_{j1})}{\sigma_j^2})x_j + \Sigma_{j=1}^n (\frac{\mu_{j1}^2 - \mu_{j0}^2}{2\sigma_j^2}) + \log\frac{p}{1-p} \leq 0$$

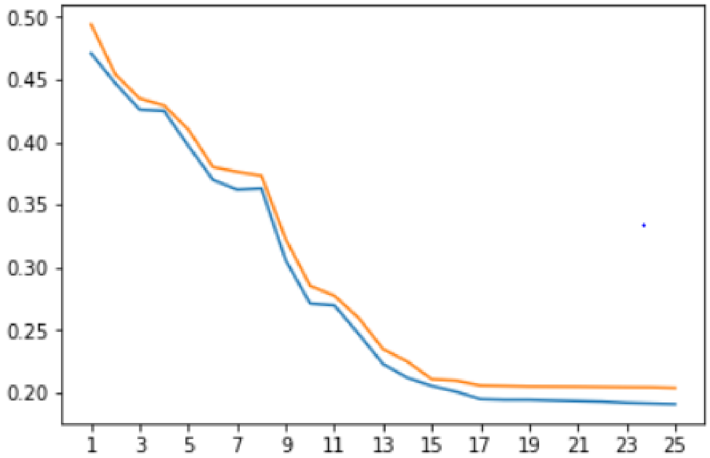
Thus we have

$$w_j = \frac{(\mu_{j0} - \mu_{j1})}{\sigma_j^2}$$

$$b = \Sigma_{j=1}^n (\frac{\mu_{j1}^2 - \mu_{j0}^2}{2\sigma_j^2}) + \log\frac{p}{1-p}$$

PROBLEM 2.

a)



Green: Training error
Yellow: Testing error

c) The model seems to stabilize around 20% and seems to never overfits after more iterations are added. It seems like using a stumps as the weak classifier reduces the chance of overfitting due to how inaccurate each individual stumps is to the model.

PROBLEM 3.

Note the LP problem is
$max_{\beta,\beta_0}\tau$, subject to $y_i(\beta^T x_i + \beta_0) \geq 1 \forall i$
Note since $\tau = \frac{1}{||\beta||}$, for β, β_0, we have $max_{\beta,\beta_0}\tau = min_{\beta,\beta_0}\frac{1}{\tau} = min_{\beta,\beta_0}||B|| = min_{\beta,\beta_0}\frac{||B||^2}{2}$
Thus, by the Lagrangian we have

$$\begin{aligned} L_p &= \frac{||\beta||^2}{2} + \Sigma_{i=1}^n \alpha_i [y_i(\beta^T x_i + \beta_0) - 1] \\ &= \frac{||\beta||^2}{2} + \Sigma_{i=1}^n \alpha_i y_i(\beta^T x_i + \beta_0) - \Sigma_{i=1}^n \alpha_i \end{aligned}$$

After setting the derivative to zero we have

$$\begin{aligned} 0 &= \beta - \Sigma_{i=1}^n \alpha_i y_i x_i \\ \beta &= \Sigma_{i=1}^n \alpha_i y_i x_i \end{aligned}$$

$$0 = \Sigma_{i=1}^n \alpha_i y_i$$

Substituting the above into the Langrangian

$$\begin{aligned} 0 &= \frac{||B||^2}{2} + \frac{1}{2}\Sigma_{i=1}^n \alpha_i y_i (\Sigma_{j=1}^n \alpha_j y_j x_j) x_i - \frac{1}{2}\alpha_i \\ \frac{||B||^2}{2} &= \frac{1}{2}\Sigma_{i=1}^n \alpha_i - \frac{1}{2}\Sigma_{i=1}^n \Sigma_{j=1}^n \alpha_i \alpha_j y_i y_j x_j x_i \\ \frac{1}{\tau^2} &= \Sigma_{i=1}^n \alpha_i - \Sigma_{i=1}^n \Sigma_{j=1}^n \alpha_i \alpha_j y_i y_j x_j x_i \end{aligned}$$

As desired