CSCI-567: Assignment #4

Due on Monday, November 9, 2015

Saket Choudhary
2170058637

Contents

Problem 1

Problem 1: (a) Gradient Calculation

$$L(y_i, \hat{y_i}) = \log(1 + \exp(-y_i\hat{y_i}))$$

$$g_i = \frac{\partial L(y_i, \hat{y_i})}{\partial \hat{y_i}}$$

$$\boxed{g_i = \frac{-y_i \exp(-y_i\hat{y_i})}{1 + \exp(-y_i\hat{y_i})}}$$

Problem 1: (b) Weak Learner Section

$$h^* = \min_{h\in H} \big(\min_{\gamma \in R} \sum_{i=1}^{n} (-g_i - \gamma h(x_i))^2\big)$$

$$\implies \frac{\partial h^*}{\partial \gamma} = 0$$

$$\implies 2\sum_{i=1}^{n}(-g_i - \gamma h(x_i))(-h(x_i)) = 0$$

$$\boxed{\hat{h} = -\frac{\sum_{i=1}^{n} g_i h(x_i)}{\sum_{i=1}^{n} h(x_i)^2}}$$

Also check if it is indeed minimum with a second derivative test:

$$\frac{\partial^2 h^*}{\partial \gamma^2} = 2\sum_{i=1}^{n} h(x_i)^2 > 0$$

Since the second derivative is positive definite, $ga\hat{m}ma$ is indeed where the minima occurs.

Problem 1 continued on next page. . .

Problem 1: (c) Step Size Selection

$$\alpha^* = \arg\min_{\alpha \in R} \sum_{1}^{n} L(y_i, \hat{y}_i + \alpha h^*(x_i))$$

Newton's approximation:

$$\alpha_1 = \alpha_0 - \frac{f(\alpha_0)}{f'(\alpha_0)}$$

We start from $\alpha_0 = 0$ and hence:

$$\begin{aligned}
f(\alpha_0) &= \sum_{i=1}^{n} \log(1 + \exp(-y_i \hat{y}_i)) \\
f'(\alpha_0) &= \sum_{i=1}^{n} \frac{\partial L}{\partial \alpha} \\
&= \sum_{i=1}^{n} \frac{-y_i h^*(x_i) \exp\left(-y_i(\hat{y}_i + \alpha h^*(x_i))\right)}{1 + \exp\left(-y_i(\hat{y}_i + \alpha h^*(x_i))\right)} \\
&= \sum_{i=1}^{n} \frac{-y_i h^*(x_i) \exp(-y_i \hat{y}_i)}{1 + \exp(-y_i \hat{y}_i)}
\end{aligned}$$

Thus,

$$\alpha_1 = \frac{\sum_{i=1}^{n} (\log(1 + \exp(-y_i \hat{y}_i)))}{\sum_{i=1}^{n} \frac{-y_i h^*(x_i) \exp(-y_i \hat{y}_i)}{1 + \exp(-y_i \hat{y}_i)}}$$

Problem 2

Problem 2: (a)

Primal form:

$$\min_w ||w||^2$$

$$\text{such that } |y_i - (w^T x_i + b)| \leq \epsilon$$

Problem 2 continued on next page...

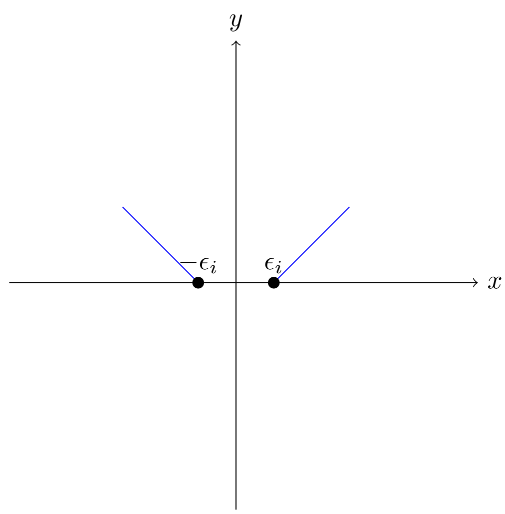
Problem 2: (b)

$$
\begin{aligned}
\min_{w,\epsilon_i} \frac{1}{2}||w||^2 + C\sum_i \epsilon_i \\
\text{such that } (w^T x_i + b) - y_i \leq n_i + \epsilon_i \text{ (positive deviation)} \\
\text{and } y_i - (w^T x_i + b) \leq p_i + \epsilon_i \text{ (negative deviation)} \\
n_i \geq 0 \\
p_i \geq 0
\end{aligned}
$$

Also, the slackness loss needs further constraints:

$$
n_i = \begin{cases} 0 & |n_i| < \epsilon_i, \\ |n_i| - \epsilon & otherwise \end{cases}
$$

$$
p_i = \begin{cases} 0 & |p_i| < \epsilon_i, \\ |p_i| - \epsilon & otherwise \end{cases}
$$



So essentially n_i, p_i are non zero, only above the two blue lines

Problem 2: (c)

$$
\begin{aligned}
L = \frac{1}{2}||w||^2 + C\sum_i (p_i + n_i) \\
- \sum_i (\eta_i p_i + \eta_i' n_i) \\
- \sum_i \alpha_i(\epsilon + p_i - (y_i - (w^T x_i + b))) \\
- \sum_i \beta_i(\epsilon + n_i + (y_i - (w^T x_i + b)))
\end{aligned}
$$

Conditions :

$$
\begin{aligned}
\alpha_i \geq 0 \\
\beta_i \geq 0 \\
\eta_i \geq 0 \\
\eta_i^* \geq 0
\end{aligned}
$$

Dual Form:

$$
\begin{aligned}
\Delta_w L &= 0 \\
&= w - \sum_i \alpha_i x_i + \sum_i \beta_i x_i = 0 \\
\Delta_b L &= 0 \\
&= \sum_i \alpha_i - \sum_i \beta_i = 0 \\
\Delta_{p_i} L &= 0 \\
&= C - \sum \eta_i - \sum_i \alpha_i = 0 \\
\Delta_{n_i} L &= 0 \\
&= C - \sum \eta_i' - \sum_i \beta_i = 0
\end{aligned}
$$

Problem 2 [Problem 2: (c)] continued on next page. . .

Thus, w is given by:

$$w = \sum_i \alpha_i x_i - \sum_i \beta_i x_i$$

depends only on the support vectors.
This reduces the optimisation to:

$$\begin{aligned}
\max f = &\frac{1}{2}\sum_{i,j}(\alpha_i - \beta_i)x_i^T x_j(\alpha_j - \beta_j) + p_i(C - \sum_i \eta_i - \sum_i \alpha_i) \\
&+ n_i(C + \sum_i \eta_i' - \sum_i \beta_i) \\
&+ \epsilon(-\sum_i \alpha_i - \sum_i \beta_i) \\
&+ \sum_i y_i(\alpha_i - \beta_i) \\
&= \frac{1}{2}\sum_{i,j}(\alpha_i - \beta_i)x_i^T x_j(\alpha_j - \beta_j) - \epsilon(\sum_i(\alpha_i + \beta_i)) \\
&+ \sum_i y_i(\alpha_i - \beta_i)
\end{aligned}$$

$$\text{such that} \sum_i (\alpha_i - \beta_i) = 0$$

$$\text{and } \alpha_i, \beta_i \in [0, C]$$

Problem 2: (d)

Using Kernel transformation:

$$w = \sum_i (\alpha_i - \beta_i)\phi(x_i)$$

this happens because $x_i' x_j$ gets mapped onto by an equivalnet kernel function $k(x_i, x_j) = \phi^T(x_i)\phi(x_j)$ and the objective function is:

$$\max_f = \frac{1}{2}\sum_{i,j}(\alpha_i - \beta_i)k(x_i, x_j)(\alpha_j - \beta_j) - \epsilon(\sum_i(\alpha_i + \beta_i))$$