Assignment

1 Regularization

We first calculate the gradient of squared loss as follows.

$$\bigtriangledown_w L_1 = \sum_{i=1}^{N} 2(y_i - w^T\phi(x_i))(-\phi(x_i)) + 2\lambda w = 0 \tag{1}$$

Using Lagrange multiplier, the second objective can be turned into following formula

$$min\{L_2\} = min\{\sum_{i=1}^{N}(y_i - w^T\phi(x_i))^2 + \alpha(\sum_{i=1}^{d} w_j^2 - \tau)\} \tag{2}$$

According to the KKT condition, the solution of $min\{L_2\}$ should satisfy the conditions

$$\bigtriangledown_w L_2 = \sum_{i=1}^{N} 2(y_i - w^T\phi(x_i))(-\phi(x_i)) + 2\alpha w = 0 \tag{3}$$

$$\alpha(\sum_{j=1}^{d} w_j^2 - \tau) = 0 \tag{4}$$

$$\alpha \geq 0 \tag{5}$$

$$\sum_{j=1}^{d} w_j^2 - \tau \leq 0 \tag{6}$$

Suppose $w_1^\star$ is the solution of 1, we now show the solution of 2 is also $w_1^\star$ in two cases of $\lambda = 0$ and $\lambda > 0$.

1. $\lambda > 0$. Let $\alpha = \lambda$, thus $w_1^\star$ is the solution of 3. Let $\tau = \sum_{j=1}^{d}(w_{1j}^\star)^2$ and equation 4, 5 and 6 are satisfied. Therefore, $w_1^\star$ is the solution of 2.

2. $\lambda = 0$. Let $\alpha = \lambda = 0$ and $w_1^\star$ is the solution of 3. Let $\tau \geq \sum_{j=1}^{d}(w_{1j}^\star)^2$ and equation 4, 5 and 6 are satisfied. Therefore, $w_1^\star$ is the solution of 2.

According to the above, the solution of equation 1 and 2 are identical if the τ(in equation 2) is chosen correctly. Therefore, the two objectives are same.

2 Naive Bayes

The Naive Bayes classifier is a function as follows:

$$\begin{aligned} h^\star(x) &= \arg\max_c \delta_c(x) \\ \delta_c(x) &= \log p(x|y=c) \end{aligned} \tag{7}$$

Suppose the feature is $x = (x_1, ..., x_n)^T$ and features are of Bernoulli distribution. $p(x|y=c) = \prod_{i=1}^{n} p_{ci}^{x_i}(1-p_{ci})^{1-x_i}$, where p_{ci} is the probability of ith item in class c(we use frequency to approximate the probability). Therefore, the corresponding Naive Bayes classifier7 can be further transformed into classifier 8.

$$\begin{aligned} h^\star(x) &= \arg\max_c \delta_c(x) \\ \delta_c(x) &= \sum_{i=1}^{n} x_i \log p_{ci} + (1-x_i) \log(1-p_{ci}) \end{aligned} \tag{8}$$