

# Machine Learning: Assignment 5

## Akshay Aradhya

November 25, 2017

### 1 Problem 1

If we have  $\text{CoX} \cap \text{CoY} \neq \emptyset$  then  $\exists$  atleast one point  $\mathbf{t} \in \text{CoX}$  and  $\mathbf{t} \in \text{CoY}$   
 $\Rightarrow \exists \alpha_i$ s and  $\beta_j$ s such that

$$t = \sum_{i=1}^N x_i \alpha_i, \alpha_i \geq 0 \text{ and } \sum_{i=1}^N \alpha_i = 1 \quad \text{and} \quad t = \sum_{j=1}^M y_j \beta_j, \beta_j \geq 0 \text{ and } \sum_{j=1}^M \beta_j = 1 \quad (1)$$

If possible let there exist  $w$  and  $w_o$  such that

$$w^T x_i + w_o > 0 \quad \forall i = 1, \dots, N \text{ and } w^T y_j + w_o < 0 \quad \forall j = 1, \dots, M$$

Now as  $\alpha_i \geq 0$  we have,

$$\alpha_i (w^T x_i + w_o) \geq 0 \quad \forall i = 1, \dots, N$$

Summing over all  $i$  we get,

$$\begin{aligned} \sum \alpha_i (w^T x_i + w_o) &> 0 \text{ as } \sum_{i=1}^N \alpha_i = 1 \\ \Rightarrow w^T \sum \alpha_i x_i + w_o \sum \alpha_i &> 0 \Rightarrow w^T t + w_o > 0 \end{aligned} \quad (2)$$

Similarly, we get

$$\sum \beta_j (w^T y_j + w_o) < 0 \Rightarrow w^T \sum \beta_j y_j + w_o \sum \beta_j < 0 \Rightarrow w^T t + w_o < 0 \quad (3)$$

Clearly, statements 2 and 3 contradict each other. Hence, there does not exist any  $w$  and  $w_o$  such that the two sets are linearly separable.

### 2 Problem 2

We have to minimize the  $E_w(w)$  function i.e. equivalent to maximizing

$$\ln \left[ \prod_{i=1}^N \sigma(w^T x_i)^{y_i} (1 - \sigma(w^T x_i))^{1-y_i} \right]$$

Now as  $0 \leq \sigma(a) \leq 1$ , we would want the  $\sigma(w^T x_i)$  and  $(1 - \sigma(w^T x_i))$  to be 1.

$$\text{for } \sigma(w^T x_i) = \frac{1}{\exp(-w^T x_i) + 1} \rightarrow 1$$

we will have  $w \rightarrow \infty$ . Hence, we get

$$\lim_{w \rightarrow \infty} \sigma(w^T x_i) = 1 \quad (4)$$

Similarly for  $(1 - \sigma(w^T x_i)) = 1 - \frac{1}{\exp(-w^T x_i) + 1} \rightarrow 1$

we will have  $w \rightarrow -\infty$ . Hence, we get

$$\lim_{w \rightarrow -\infty} 1 - \sigma(w^T x_i) = 1 \quad (5)$$

Therefore, we have the required result. Also, to penalize the high magnitudes of  $w$ , we add an error term like in the ridge regression.

$$E(w) = -\ln \left[ \prod_{i=1}^N \sigma(w^T x_i)^{y_i} (1 - \sigma(w^T x_i))^{1-y_i} \right] + \lambda \|w\|_q^2$$

### 3 Problem 3

We could use the basis function  $\Phi(x_1, x_2) = \frac{x_2}{x_1}$ . This would make the circles transform to positive values and crosses to negative values.

### 4 Problem 4

The two points on the hyperplane are (0,5) and (2,0). As the data is two-dimensional, the hyperplane would be a line. Let the equation of the line be:

$$w_0 + w_1 x_1 + w_2 x_2 = 0$$

On replacing the values (0,5) and (2,0) we get,

$$w_0 = -5w_2 \text{ and } w_0 = -2w_1 \Rightarrow w_1 = \frac{5}{2}w_2$$

Taking  $w_2 = 1$ , the equation of the line is:

$$-5 + \frac{5}{2}x_1 + x_2 = 0$$