## UML 15.1

1. Show that the hard-SVM rule, namely,

$$\underset{(\mathbf{w},b):\|\mathbf{w}\|=1}{\operatorname{argmax}} \quad \min_{i\in[m]} |\langle \mathbf{w}, \mathbf{x}_i\rangle + b| \quad \text{s.t.} \quad \forall i,\ y_i(\langle \mathbf{w}, \mathbf{x}_i\rangle + b) > 0,$$

is equivalent to the following formulation:

$$\underset{(\mathbf{w},b):\|\mathbf{w}\|=1}{\operatorname{argmax}} \quad \min_{i\in[m]} y_i(\langle \mathbf{w}, \mathbf{x}_i\rangle + b).$$

First of all, we define that $G = \{(\boldsymbol{w}, b) : \forall i, y_i(\langle \boldsymbol{w}, x_i\rangle + b) > 0\}$. In this definition, $y_i$ can be either 1 or -1. With this in consideration, we can know that

$$y_i(\langle \boldsymbol{w}, x_i\rangle + b) = |\langle \boldsymbol{w}, x_i\rangle + b|, \forall(\boldsymbol{w}, b) \in G$$

Further, we can get

$$min_{i\in[m]} y_i(\langle \boldsymbol{w}, x_i\rangle + b) = min_{i\in[m]} |\langle \boldsymbol{w}, x_i\rangle + b|, \forall(\boldsymbol{w}, b) \in G \quad (1)$$

For the hard-SVM situation, all the points are linearly separable, which means that there exists a half space (**w**, b) such that

$$y_i = sign(\langle \boldsymbol{w}, x_i\rangle + b), \forall i \in [m]$$

Suppose we have found the$i = argmin_{i\in[m]} y_i(\langle \boldsymbol{w}, x_i\rangle + b)$ , since the data is linearly separable, we can find out the corresponding (**w**, b) (we don't exactly know the values) for this specific i such that

$$\{(\boldsymbol{w}, b) | y_i = sign(\langle \boldsymbol{w}, x_i\rangle + b), i = argmin_{i\in[m]} y_i(\langle \boldsymbol{w}, x_i\rangle + b)\} \quad (2)$$

Then, obviously for the (**w**, b) pairs in (1), they satisfies the following inequality,

$$y_i(\langle \boldsymbol{w}, x_i\rangle + b) > 0$$

So the half space (**w**, b) that satisfies (1) gives (2) as shown in the following

$$(\boldsymbol{w}, b) = argmax_{(\boldsymbol{w},b):\|\boldsymbol{w}\|=1} min_{i\in[m]} y_i(\langle \boldsymbol{w}, x_i\rangle + b) \in G \quad (3)$$

With (1) and (3) known to us, we prove the following equation required:

$$argmax_{(\boldsymbol{w},b):\|\boldsymbol{w}\|=1} min_{i\in[m]} y_i(\langle \boldsymbol{w}, x_i\rangle + b)$$

is equivalent to

$$argmax_{(\boldsymbol{w},b):\|\boldsymbol{w}\|=1} min_{i\in[m]} |\langle \boldsymbol{w}, x_i\rangle + b| \; s.t. \; \forall i, y_i(\langle \boldsymbol{w}, x_i\rangle + b) > 0$$

## UML 15.4

Firstly, we know that

$$min_{x\in X}f(x,y) \leq \ f(x,y), \forall x \in X$$

We get max over y on the two sides,

$$max_{y\in Y}min_{x\in X}f(x,y) \leq max_{y\in Y}f(x,y), \forall x \in X, y \in Y \quad (4)$$

$\forall x \in X, y \in Y$, we have equation (4), then obviously, we have

$$max_{y\in Y}min_{x\in X}f(x,y) \leq min_{x\in X}max_{y\in Y}f(x,y), \forall x \in X, y \in Y \quad (4)$$

## BRML 11.9

$$\theta = \begin{bmatrix} 0.3 & 0.3 \\ 0.2 & 0.2 \end{bmatrix}$$

$$\theta = \begin{bmatrix} 0.2 & 0.4 \\ 0.4 & 0 \end{bmatrix}$$

When the row values are added, we can see that

$$P(x_1) = 0.6, P(x_2) = 0.4$$

They share the same marginal likelihood solution

## BRML 20.1

Suppose the $v_3^5$ is unknown,

$$P(\boldsymbol{V}) = \sum_h P(h)\big(P(v_1|h)P(v_2|h)P(v_4|h)P(v_5|h)\big)$$

When we take logarithmic values on the both sides, we get

$$\frac{\partial P(\boldsymbol{V})}{\partial p(v_i|h)} = \sum_h P(h) \sum \log P(v_1|h) + \log P(v_2|h) + \log P(v_4|h) + \log P(v_5|h)$$

## BRML 20.4

$$L = \sum_{n=1}^{N} \langle \log p(h) \rangle_{p^{old}(h|v^n)} + \lambda\left(1 - \sum_h p(h)\right) = \sum_{n=1}^{N}\sum_h \log p(h) \cdot p^{old}(h|v^n) + \lambda\left(1 - \sum_h p(h)\right)$$

$$= \sum_h \log p(h) \sum_{n=1}^{N} p^{old}(h|v^n) + \lambda\left(1 - \sum_h p(h)\right)$$

$$\frac{\partial L}{\partial p(h)} = 0, \frac{\sum_{n=1}^{N} p^{old}(h|v^n)}{p(h)} = \lambda, p(h) = \frac{\sum_{n=1}^{N} p^{old}(h|v^n)}{\lambda}$$

$$\sum_h p(h) = 1, \qquad \frac{\sum_h \sum_{n=1}^{N} p^{old}(h|v^n)}{\lambda} = 1, \;\; \lambda = N$$

So, we have

$$p(h) = \frac{\sum_{n=1}^{N} p^{old}(h|v^n)}{N}$$