Machine Learning HW2

Problem 1

Solution:

For the linear SVM in the non-separable case, we are trying to optimize following equation:

$\min_{w,b,\epsilon_i} \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^N \epsilon_i$

s.t. $y_i(\langle w, x_i\rangle + b) \geq 1 - \epsilon_i, \epsilon_i \geq 0, \forall i = 1, \dots, N$

Using Lagrange multiplier method and treat it as dual problem, we are trying to optimize following equation (denote this equation as equation (*)):

$\max_{\alpha,\beta} \min_{w,b,\epsilon_i} L(w, b, \epsilon, \alpha, \beta)$
$= \max_{\alpha,\beta} \min_{w,b,\epsilon_i} \frac{1}{2}w^T w + C\sum_{i=1}^N \epsilon_i - \sum_{i=1}^N \alpha_i[y_i(\langle w, x_i\rangle + b) - 1 + \epsilon_i] - \sum_{i=1}^N \beta_i\epsilon_i \ (*)$

Calculating partical derivative, we can get:

$\frac{\partial L}{\partial w} = w - \sum_{i=1}^N \alpha_i y_i x_i = 0$

$\frac{\partial L}{\partial b} = -\sum_{i=1}^N \alpha_i y_i = 0$

$\frac{\partial L}{\partial \epsilon_i} = C - \alpha_i - \beta_i = 0, \forall i = 1, \dots, N$

That is, we get following restrictions:

$w = \sum_{i=1}^N \alpha_i y_i x_i$

$\sum_{i=1}^N \alpha_i y_i = 0$

$C = \alpha_i + \beta_i, \forall i = 1, \dots, N$

Considering these restrictions in equation (*), we are now trying to optimize following equation:

$\max_\alpha \sum_{i=1}^N \alpha_i - \frac{1}{2}\sum_{i=1}^N \sum_{j=1}^N y_i y_j \alpha_i \alpha_j \langle x_i, x_j\rangle$

s.t. $\sum_{i=1}^N \alpha_i y_i = 0, 0 \leq \alpha_i \leq C, \forall i = 1, \dots, N$

Assuming the solution to this optimization problem is α^*, we get the solution:

$w^* = \sum_{i=1}^N y_i \alpha_i^* x_i = \sum_{\alpha_i^* \neq 0} y_i \alpha_i^* x_i$

$b^* = y_i - \langle w^*, x_i\rangle = y_i - \sum_{\alpha_i^* \neq 0} y_i \alpha_i^* \langle x_i, x_j\rangle$

Finally, we get the hyperplane:

$f(x) = \langle w^*, x\rangle + b^* = \sum_{\alpha_i^* \neq 0} y_i \alpha_i^* \langle x_i, x\rangle + b^*$

Problem 2

Solution:

- Non-negativity

Since euclidean distance is always no less than 0, $h(A,B)=\max_{a\in A}\min_{b\in B}||a-b||_2\ge 0$, and $h(B,A)=\max_{b\in B}\min_{a\in A}||b-a||_2\ge 0$, so $H(A,B)=\max(h(A,B),h(B,A))\ge 0$.

- Identity

Supposing $A=B$, we can easily get $\forall a\in A$, $\exists b=a\in B$, $||a-b||_2=0$, and $\forall b\in B$, $\exists a=b\in A$, $||a-b||_2=0$. That is $h(A,B)=h(B,A)=0$. So $H(A,B)=0$.

Supposing $H(A,B)=0$, we can easily get $h(A,B)=h(B,A)=0$. That is, $\forall a\in A$, $\min_{b\in B}||a-b||_2=0$, and $\forall b\in B$, $\min_{a\in A}||a-b||_2=0$. Supposing $\exists a_0\in A$, $a_0\notin B$, we can easily get $\forall b\in B$, $||a_0-b||_2>0$, we get a contradiction. So $A=B$.

- Symmetry

$H(A,B)=\max(h(A,B),h(B,A))=\max(h(B,A),h(A,B))=H(B,A)$

- Triangle inequality

Supposing A,B,C are bounded closed set, $b^*\in B$ satisfy $h(A,B)$, then we can get:

$$
\begin{aligned}
&h(A,B)+h(B,C)\\
&=\max_{a\in A}\min_{b\in B}||a-b||_2+\max_{b\in B}\min_{c\in C}||b-c||_2\\
&=\max_{a\in A}||a-b^*||_2+\max_{b\in B}\min_{c\in C}||b-c||_2\\
&\ge\max_{a\in A}||a-b^*||_2+\min_{c\in C}||b^*-c||_2\\
&=\max_{a\in A}\min_{c\in C}(||a-b^*||_2+||b^*-c||_2)\\
&\ge\max_{a\in A}\min_{c\in C}||a-c||_2\\
&=h(A,C)
\end{aligned}
$$

Similarly, we can get:

$$h(B,A)+h(C,B)\ge h(C,A)$$

That is to say:

$$
\begin{aligned}
&H(A,B)+H(B,C)\\
&=\max(h(A,B),h(B,A))+\max(h(B,C),h(C,B))\\
&\ge\max(h(A,C),h(C,A))\\
&=H(A,C)
\end{aligned}
$$

Problem 3

Solution:

1. Supposing we observe nothing, since a,b are indepentent to each other, we can get the following equation:

   $p(a,b)=p(a,b)\sum_d\sum_c p(c|a,b)p(d|c)=p(a)p(b)\sum_d p(d|a,b)=p(a)p(b)$

   So we can say:

   $a\perp\!\!\!\perp b|\phi$

2. Supposing we now obeserve d, we can get the following equations:

$$p(a,b|d) = \frac{p(a,b,d)}{p(d)} = p(a)p(b)\frac{\sum_c p(c|a,b)p(d|c)}{p(d)} = \frac{p(a)p(b)p(d|a,b)}{p(d)}$$

$$p(a|d) = \frac{p(a,d)}{p(d)} = p(a)\frac{\sum_b p(b)\sum_c p(c|a,b)p(d|c)}{p(d)}$$

$$p(b|d) = \frac{p(b,d)}{p(d)} = p(b)\frac{\sum_a p(a)\sum_c p(c|a,b)p(d|c)}{p(d)}$$

Since:

$$p(a,b|d) \neq p(a|d)p(b|d)$$

So we can say:

$$a \not\perp\!\!\!\perp b|d$$