| | |
|---|---|
| **CS446: Machine Learning** | **Fall 2015** |

Problem Set 5

Jifu Zhao — *Handed In: 11/06/2015*

1. **Answer to problem 1**

(a) **1,** $\mathbf{W} = [-1, 1]^T$
$\theta = 0$

**2,** $\mathbf{W} = [-0.5, 0.25]^T$
$\theta = 0$

**3,** For SVM method in this problem, we need to find three lines that

$$w_1 x1 + w_2 x_2 + \theta = 0$$

$$w_1 x1 + w_2 x_2 + \theta = -1$$

$$w_1 x1 + w_2 x_2 + \theta = 1$$

The first step is to find the two support vectors. Considering all the points from the table, it is easy to conclude that **(-1.2, 1.6, (+1))** and **(2, 0, (-1))** has the least distance. So, we can derive:

$$-1.2w_1 + 1.6w_2 x_2 + \theta = 1$$

$$2w_1 + 0w_2 + \theta = -1$$

Through these two questions, we can get that: $\theta = -2w_1 - 1$ and $w_2 = 2w_1 + 1.25$. So, we can derive that

$$w_1^2 + w_2^2 = 5w_1^2 + 5w_1 + 1.25^2$$

so, in order to get the least value for $w_1^2 + w_2^2$, we need $\mathbf{w_1} = \mathbf{-0.5}$, so $\mathbf{w_2} = \mathbf{0.25}$ and $\theta = \mathbf{0}$.

So, we need that: $\mathbf{W} = [\mathbf{-0.5}, \mathbf{0.25}]^T$, $\qquad \theta = \mathbf{0}$.

(b) **1,** $\mathbf{I} = \{\mathbf{1}, \mathbf{6}\}$ or $\mathbf{I} = \{(\mathbf{-1.2}, \mathbf{1.6}), (\mathbf{2}, \mathbf{0})\}$

**2,** $\alpha = \{\mathbf{\frac{5}{32}}, \mathbf{\frac{5}{32}}\}$

**3,** $objective function value = \mathbf{\frac{5}{32}}$

**(c)** C controls the tradeoff between large margin (small $||w||$) and small loss. When C is larger, we concerned more on the mistakes. When C is smaller, we concerned more on larger margin.

**1**, When $C = \infty$, we need $\xi_i$ to be 0, so, when $C = \infty$, we get the same answer in (a)-2.
**2**, When $C = 1$, we will make more misclassifications, but we can also have a larger margin.
**3**, When $C = 0$, we will concerned more on the larger margin. So, we have larger margin, but we will also make more misclassifications.

2. **Answer to problem 2**

**(a) Dual Perceptron:**

**1,** Initialize $\alpha$ and $\theta$ to zero vectors
**2,** For each example $(x_m, y_m)$:
if $y_m \sum_{i=1}^{m} \alpha_i x_i x_m < 0$, $\alpha_m = \alpha_m + 1$
**3,** Output the final result $w$

**(b)** Suppose that: $\vec{x} = (x_1, x_2)^T$ and $\vec{z} = (z_1, z_2)^T$, so

$$K(\vec{x}, \vec{z}) = (\vec{x}^T \vec{z})^3 + 400(\vec{x}^T \vec{z})^2 + 100(\vec{x}^T \vec{z})$$

can be written as:

$$K(\vec{x}, \vec{z}) = (x_1 z_1 + x_2 z_2)^3 + 400(x_1 z_1 + x_2 z_2)^2 + 100(x_1 z_1 + x_2 z_2)$$

**First**, let's prove that $K_1 = (x_1 z_1 + x_2 z_2)^3$ is a valid kernel.

$$K_1 = x_1^3 z_1^3 + 3x_1^2 x_2 z_1^2 z_2 + 3x_1 x_2^2 z_1 z_2^2 + x_2^3 z_2^3$$

$$K_1 = (x_1^3, \sqrt{3}x_1^2 x_2, \sqrt{3}x_1 x_2^2, x_2^3)(z_1^3, \sqrt{3}z_1^2 z_2, \sqrt{3}z_1 z_2^2, z_2^3)^T = \psi_1(\vec{x})^T \psi_1(\vec{z})$$

**So, $K_1$ is a valid kernel.**

**Now**, let's prove that $K_2 = 400(x_1 z_1 + x_2 z_2)^2$ is a valid kernel.

$$K_2 = 400(x_1^2 z_1^2 + 2x_1 x_2 z_1 z_2 + x_2^2 z_2^2)$$

$$K_2 = 20(x_1^2, \sqrt{2}x_1 x_2, x_2^2)20(z_1^2, \sqrt{2}z_1 z_2, z_2^2)^T = \psi_2(\vec{x})^T \psi_2(\vec{z})$$

**So, $K_2$ is valid kernel.**

**Finally**, $K_3 = 100(x_1 z_1 + x_2 z_2) = 10(x_1, x_2)10(z_1, z_2)^T = \psi_3(\vec{x})^T \psi_3(\vec{z})$ is also a kernel.

Since, $K_1$, $K_2$ and $K_3$ are all valid kernels, so

$$K(\vec{x},\vec{z}) = K_1 + K_2 + K_3 = \psi_1(\vec{x})^T\psi_1(\vec{z}) + \psi_2(\vec{x})^T\psi_2(\vec{z}) + \psi_3(\vec{x})^T\psi_3(\vec{z})$$

$$K(\vec{x},\vec{z}) = [\psi_1(\vec{x})\psi_2(\vec{x})\psi_3(\vec{x})][\psi_1(\vec{z})\psi_2(\vec{z})\psi_3(\vec{z})]^T$$

So, $\mathbf{K}(\tilde{\mathbf{x}},\tilde{\mathbf{z}})$ is also valid kernels.

3. **Answer to problem 3**

| | | Hypothesis 1 | | | | Hypothesis 2 | | | |
|---|---|---|---|---|---|---|---|---|---|
| $i$ | Label | $D_0$ | $x_1 \equiv$ $[\mathbf{x} > \mathbf{5}]$ | $x_2 \equiv$ $[\mathbf{y} > \mathbf{6}]$ | $h_1 \equiv$ $[\mathbf{x_1}]$ | $D_1$ | $x_1 \equiv$ $[\mathbf{x} > \mathbf{3}]$ | $x_2 \equiv$ $[\mathbf{y} > \mathbf{8}]$ | $h_2 \equiv$ $[\mathbf{x_2}]$ |
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| *1* | − | 1/10 | − | + | − | 1/16 | − | + | + |
| *2* | − | 1/10 | − | − | − | 1/16 | + | − | − |
| *3* | + | 1/10 | + | + | + | 1/16 | + | − | − |
| *4* | − | 1/10 | − | − | − | 1/16 | + | − | − |
| *5* | − | 1/10 | − | + | − | 1/16 | − | + | + |
| *6* | + | 1/10 | + | + | + | 1/16 | + | − | − |
| *7* | + | 1/10 | + | + | + | 1/16 | + | + | + |
| *8* | − | 1/10 | − | − | − | 1/16 | + | − | − |
| *9* | + | 1/10 | − | + | − | 1/4 | + | + | + |
| *10* | − | 1/10 | + | + | + | 1/4 | + | − | − |

Table 1: Table for Boosting results

**(a)** There are 10 samples in total, so $m = 10$, in this way, $D_0 = 1/m = 0.1$.
In order to reduce the error rate, through comparison, finally we find that for x and y, when we choose $x > 5$ and $y > 6$, the error rate will be smallest, which are 0.2 and 0.3. So, $\mathbf{x_1} = [\mathbf{x} > \mathbf{5}]$ and $\mathbf{x_2} = [\mathbf{y} > \mathbf{6}]$.

**(b)** Since error rate is 0.2 and 0.3, we should choose $\mathbf{x_1} = [\mathbf{x} > \mathbf{5}]$ as the first hypothesis $\mathbf{h_1} = [\mathbf{x_1}]$. The prediction are shown in tables.

**(c)** From (a), we know that error rate $\epsilon = 0.2$. According to

$$\alpha = 0.5 * ln[(1-\epsilon)/\epsilon]$$

we can get that $\alpha = ln2 = 0.693$. According to

$$z_t = \sum D_t exp(-\alpha_t y h(x))$$

When $y_i = h(x_i)$, $D_t exp(-\alpha_t y h(x)) = 1/10 * exp(-ln2) = 0.05$,
When $y_i \neq h(x_i)$, $D_t exp(-\alpha_t y h(x)) = 1/10 * exp(ln2) = 0.2$.

So,

$$z_t = \sum D_t exp(-\alpha_t y h(x)) = 8 * 0.05 + 2 * 0.2 = 0.8$$

Then, according to $D_{t+1} = D_t / z_t * exp(-\alpha_t y h(x))$,

**when** $y_i = h(x_i)$**,**

$$D_1 = 0.1/0.8 * exp(-ln2) = 1/16$$

**when** $y_i \neq h(x_i)$**,**

$$D_1 = 0.1/0.8 * exp(ln2) = 1/4$$

So, we can fill in the Table as shown above.

**(d)** First, let's calculate $\alpha_2$. For $h_2 = x_2 = [y > 8]$, error rate $\epsilon = 4/16 = 1/4$, so, according to

$$\alpha = 0.5 * ln[(1 - \epsilon)/\epsilon]$$

we can get that

$$\alpha_2 = 0.5 * ln[(1 - 0.25)/0.25] = 1/2 * ln3 = 0.549$$

So, $\alpha_{\mathbf{1}} = \mathbf{0.693}$, $\alpha_{\mathbf{2}} = \mathbf{0.549}$

According to the above analysis, we can write the final hypothesis as follows:

$$\mathbf{h} = \mathbf{sign}(\sum \alpha_{\mathbf{t}} \mathbf{h_t}) = \mathbf{sign}[\mathbf{0.693}(\mathbf{x} > \mathbf{5}) + \mathbf{0.549}(\mathbf{y} > \mathbf{8})]$$

4. **Answer to problem 4**

**(a) i. The expected number of children in towns A is 1 and 2 in towns B**

**Proof:**
In **town A**, each family has just one child, so the expected number of children in towns A is 1.

In **town B**, the probability of a family has k children is:

$$P(k) = (\frac{1}{2})^{k-1}\frac{1}{2} = (\frac{1}{2})^k$$

so the expected number of children will be:

$$N = \sum_1^\infty k(\frac{1}{2})^k = 2$$

So, the expected number of children in towns B is **2**.

**ii. The boy to girl ratio in towns A is 1 : 1 and 1 : 1 in towns B**

**Proof:**
In **town A**, at first the boy to girl ratio is 1:1. So, the probability of boy and girl is 0.5 and 0.5. Now, when they have children, they will have only one child, with the probability of 0.5 to be a boy and 0.5 to be a girl. So, at the end of the first generation, the probability of boy will be:$0.5 * 0.5 * 2 = 0.5$ and the probability of girl will be:$0.5 * 0.5 * 2 = 0.5$. So, the boy to girl ratio in towns A is $\mathbf{1:1}$.

In **town B**, at first the boy to girl ratio is 1:1. So, the probability of boy and girl is 0.5 and 0.5. Now, for each family, there be definitely one boy, so the number of boy for each family will be: 1. For girls, the expected number of girls will be:

$$\sum_{1}^{\infty} k(\frac{1}{2})^{k+1} = 1$$

So, after one generation, the expected number of boys will be: $0.5 * 1 = 0.5$. and the expected number of girls will be: $0.5 * 1 = 0.5$. So the boy to girl ratio in towns B is $\mathbf{1:1}$.

**(b) i. Proof:**

$$P(A|B) = \frac{P(A,B)}{P(B)}$$

Since $P(A,B) = P(A)P(B|A)$, so

$$\frac{P(A,B)}{P(B)} = \frac{P(A)P(B|A)}{P(B)}$$

So,

$$\mathbf{P(A|B) = \frac{P(B|A)P(A)}{P(B)}}$$

**ii.**

$$P(A,B,C) = P(A|B,C)P(B,C)$$

Since

$$P(B,C) = P(B|C)P(C)$$

So:

$$P(A,B,C) = P(A|B,C)P(B,C) = P(A|B,C)P(B|C)P(C)$$

So:

$$\mathbf{P(A,B,C) = P(A|B,C)P(B|C)P(C)}$$

**(c) Proof:**
First, let's calculate the probability distribution of X. Suppose that for n events, A events will occur for m times. So,

$$P(A) = \frac{m}{n}$$

And, it is clear that:

$$P(X = 1) = \frac{m}{n} \qquad P(X = 0) = \frac{n - m}{n}$$

So,

$$E(X) = P(X = 1) * 1 + P(X = 0) * 0 = P(X = 1) = \frac{m}{n}$$

So,

$$\mathbf{E}(\mathbf{X}) = \mathbf{P}(\mathbf{A})$$

**(d)** First, let's calculate the distribution for (X, Y). The result are shown in following Table 2:

| | $X = 0$ | $X = 1$ | |
|---|---|---|---|
| $Y = 0$ | 30/90 | 18/90 | 48/90 |
| $Y = 1$ | 25/90 | 17/90 | 42/90 |
| | 55/90 | 35/90 | 1 |

Table 2: T(X, Y) Distribution

**i. X is not independent of Y.**

**Reason**: If X is independent of Y, we should have: $P(X, Y) = P(X)P(Y)$. So, for example, from Table 2, we can know that:$P(X = 1, Y = 1) = 18/90$, and $P(X = 1) = 35/90$, $P(Y = 1) = 42/90$. It is clear that $P(X, Y) \neq P(X)P(Y)$.

So, X is not independent of Y.

**ii. X is conditionally independent of Y given Z**

**Reason**: According to Table 3:

| | $Z = 0$ | | $Z = 1$ | |
|---|---|---|---|---|
| | $X = 0$ | $X = 1$ | $X = 0$ | $X = 1$ |
| $Y = 0$ | 1/15 | 1/15 | 4/15 | 2/15 |
| $Y = 1$ | 1/10 | 1/10 | 8/45 | 4/45 |

Table 3: Table for (X, Y) Distribution

We can get that:
$P(Z=0)=30/90$, $P(X=0|Z=0)=15/30$, $P(X=1|Z=0)=15/30$, $P(Y=0|Z=0)=12/30$, $P(Y=1|Z=0)=18/30$,

So,

$$\mathbf{P(X,Y|Z=0)=P(X|Z=0)P(Y|Z=0)}$$

Similarly,
$P(Z=1)=60/90$, $P(X=0|Z=1)=40/60$, $P(X=1|Z=1)=20/60$, $P(Y=0|Z=1)=36/60$, $P(Y=1|Z=1)=24/60$,
So,

$$\mathbf{P(X,Y|Z=1)=P(X|Z=1)P(Y|Z=1)}$$

In conclusion, **X is conditionally independent of Y given Z**.

**iii.** From Table 2, we can know that:

$$\{X+Y>0\}=\{(X=0,Y=1),(X=1,Y=0),(X=1,Y=1)\}$$

So,

$$P(X=0|X+Y>0)=P(X=0|\{(X=0,Y=1),(X=1,Y=0),(X=1,Y=1)\})$$

So

$$P(X=0|X+Y>0)=\frac{P(X=0,X+Y>0)}{P(\{(X=0,Y=1),(X=1,Y=0),(X=1,Y=1)\})}$$

And

$$P(X=0,X+Y>0)=P(X=0,Y=1)=25/90$$

$$P(\{(X=0,Y=1),(X=1,Y=0),(X=1,Y=1)\})=25/90+18/90+17/90=60/90$$

So

$$P(X=0|X+Y>0)=\frac{25/90}{60/90}=5/12$$

So,

$$\mathbf{P(X=0|X+Y>0)=5/12}$$