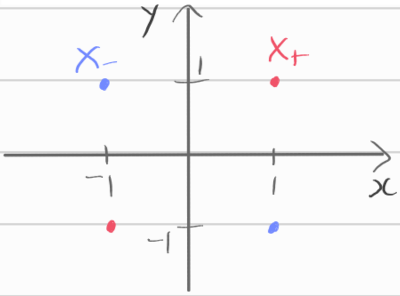
2018-17119 박수현 컴퓨터비전 HW5 Theory

1.1 SVM (20 points)

Suppose that training examples are points in ~~3-D~~ 2-D space. The positive examples are $X_+ = \{(1,1), (-1,-1)\}$. The negative examples are $X_- = \{(1,-1), (-1,1)\}$.

(a) [4 pts] Are the positive examples linearly separable from the negative examples? (*i.e.* can you draw a line to separate the positive examples from negative examples)?



No.

As seen in the left graph, a single line cannot separate positive examples from the negative examples. Linear boundary fails.

(b) [8 pts] Consider the feature transformation $\phi(x) = [1, x, y, xy]^T$, where x and y are the first and second coordinates of an example. Write down the transformed coordinates of X_+ and X_- (*i.e.* $\phi(X_+)$ and $\phi(X_-)$ for all four examples).

For $(x,y) \in X_+$, $\phi((1,1)) = [1,1,1,1]^T$, $\phi((-1,-1)) = [1,-1,-1,1]^T$

For $(x,y) \in X_-$, $\phi((1,-1)) = [1,1,-1,-1]^T$, $\phi((-1,1)) = [1,-1,1,-1]^T$

$\therefore \phi(X_+) = \{(1,1,1,1), (1,-1,-1,1)\}$, $\phi(X_-) = \{(1,1,-1,-1), (1,-1,1,-1)\}$

(c) [8 pts] Consider the prediction function $y(x) = w^T\phi(x)$. Give the coefficient w of a maximum-margin decision surface separating the positive from the negative examples.

(hint: w is $[4 \times 1]$ vector, whose elements are only 0 or 1).

Let $w = [w_1, w_2, w_3, w_4]^T$, $w_i = 0$ or 1

$\rightarrow \min_w \|w\|^2$ s.t. $y_i(w^T\phi(x_i)) \geq 1$ for $i = 1, 2, 3, 4$

To express this problem as a maximization problem (dual form):

$\rightarrow \max_\alpha \sum_{i=1}^{4} \alpha_i - \frac{1}{2}\sum_{i=1}^{4}\sum_{j=1}^{4} \alpha_i\alpha_j y_i y_j \phi(x_i)^T\phi(x_j)$ s.t. $0 \leq \alpha_i \leq C$ for all i and $\sum_{i=1}^{N}\alpha_i y_i = 0$

$\rightarrow \max_\alpha \sum_{i=1}^{4} \alpha_i - \frac{1}{2}(4\alpha_1\alpha_1 + 4\alpha_2\alpha_2 + 4\alpha_3\alpha_3 + 4\alpha_4\alpha_4)$

$\rightarrow \max_\alpha \alpha_1(1-2\alpha_1) + \alpha_2(1-2\alpha_2) + \alpha_3(1-2\alpha_3) + \alpha_4(1-2\alpha_4)$

Since constraint: $\sum_{i=1}^{N} a_i y_i = \alpha_1 - \alpha_2 - \alpha_3 + \alpha_4 = 0$

$\rightarrow \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \frac{1}{4}$

$w = \sum_{i=1}^{4} \alpha_i y_i \phi(x_i) = \frac{1}{4}\cdot 1 \cdot [1,1,1,1]^T + \frac{1}{4}\cdot 1\cdot[1,-1,-1,1]^T + \frac{1}{4}\cdot(-1)\cdot[1,1,-1,-1]^T + \frac{1}{4}\cdot(-1)\cdot[1,-1,1,-1]^T$

$= [0,0,0,1]^T$

1.2 K-means (20 points)

(a) [4 pts] Mahalanobis measure is one of many distance measure used for k-means. Given the below definition, describe the shape of the covariance matrix corresponding to σ_i and explain why Mahalanobis measure is also called scaled euclidean measure in this case.

$$d(x,c) = \sqrt{\sum_{i=1} \frac{(x_i - c_i)^2}{s_i^2}} \quad \cdots \quad ①$$

As Mahalanobis measure is a measure of the distance between a point P and a distribution D, it takes into account the correlation between data. The formal definition of the Mahalanobis distance is as follows: $d(x,c) = \sqrt{(x-c)^T C^{-1} (x-c)}$, C is the covariance matrix $\cdots ②$

We must find C^- that satisfies ② = ①. This equation can be expressed by matrices:

$$\sqrt{\begin{bmatrix} x_1-c_1 & x_2-c_2 & \cdots & x_n-c_n \end{bmatrix} C^{-1} \begin{bmatrix} x_1-c_1 \\ x_2-c_2 \\ \vdots \\ x_n-c_n \end{bmatrix}} = \sqrt{\sum_{i=1} \frac{(x_i-c_i)^2}{\sigma_i^2}}$$

Therefore, C^{-1} should be $\begin{bmatrix} 1/\sigma_1^2 & 0 & \cdots & 0 \\ 0 & 1/\sigma_2^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 1/\sigma_n^2 \end{bmatrix}$

$$(C^{-1})^{-1} = C = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_n^2 \end{bmatrix} \leftarrow \text{covariances are all 0s}$$

$\therefore$ The covariance matrix is a diagonal matrix with σ_i as the diagonal elements, i.e., $C = \{\sigma_1^2, \cdots, \sigma_n^2\}$

The Euclidean distance is defined as follows: $d(x,c) = \sqrt{\sum_{i=1} (x_i - c_i)^2}$

In the formal Euclidean distance, since distance between every data are added, the distance value can be flawed due to different scales and correlation between data. Now x can be standardized into $z = \frac{x-c}{\sigma}$, where c is the vector of mean values.

Thus, the scaled Euclidean measure formula becomes $d(x,c) = \sqrt{\sum_{i=1} \frac{(x_i-c_i)^2}{\sigma_i^2}}$, which is equal to the given Mahalanobis formula.

(b) [8 pts] Given a dataset $\mathcal{X} = \{0, 2, 4, 6, 18, 20\}$, initialize the k-means clustering algorithm with 2 cluster centers $c_1 = 3$ and $c_2 = 4$. What are the values of c_1 and c_2 after the first iteration of k-means? Also report the values after the second iteration.

As the dataset is one-dimensional, we use the Euclidean as the distance measure.

$d(p, q) = |p - q|$

x	0	2	4	6	18	20
distance from $c_1 = 3$	3	1	1	3	15	17
distance from $c_2 = 4$	4	2	0	2	14	16
distance from $c_1 = 1$	1	1	3	5	17	19
distance from $c_2 = 12$	12	10	8	6	6	8

after first iteration:

$\Rightarrow \frac{0+2}{2} = 1 \quad \therefore c_1 = 1$

$\frac{4+6+18+20}{4} = 12 \quad c_2 = 12$

after second iteration:

$\Rightarrow \frac{0+2+4+6}{4} = 3 \quad \therefore c_1 = 3$

$\frac{18+20}{2} = 19 \quad c_2 = 19$

(c) [8 pts] Given the same dataset as in (b), perform greedy initialization to get the initial $k = 3$ centers. Start with $c_1 = 4$. Below is the greedy initialization process.

1. Choose c_1
2. Choose the next center c_i to be $\text{argmax}_{x \in \mathcal{X}}\{D(X)\}$
3. Repeat step 2 until k centers are chosen

where at any given time, with the current set of cluster centers $\mathcal{C}$,

$D(x) = \min_{c \in \mathcal{C}} ||x - c||_2$ $\quad X = (0, 2, 4, 6, 18, 20)$

$i = 1$: $c_1 = 4$

$C = \{4\}$

$i = 2$: $C = \{4, 20\}$

x	0	2	4	6	18	20	
$D(x)_1$	4	2	0	2	14	16	$\Rightarrow c_2 = 20$
$D(x)_2$	$< \overset{4}{\cancel{20}}$	$< \overset{2}{\cancel{18}}$	$< \overset{0}{\cancel{16}}$	$< \overset{2}{\cancel{14}}$	$< \overset{\cancel{14}}{2}$	$< \overset{\cancel{16}}{0}$	$\Rightarrow c_3 = 0$

$\therefore C = \{0, 4, 20\}$

cf) In a more mathematical way, let the step of iteration be i.

$i = 0$: $c_1 = 4 \quad (k=1)$

$i = 1$: Choose the next center c_2 to be $\underset{x \in X}{\text{argmax}}\{D(X)\}$

$D(0) = 4,\ D(2) = 2,\ D(4) = 0,\ D(6) = 2,\ D(18) = 14,\ D(20) = 16$

$\therefore c_2 = 20 \quad (k=2)$

$i = 2$: Choose the next center c_3 to be $\underset{x \in X}{\text{argmax}}\{D(X)\}$

$D(0) = \min(|0-4|, |0-20|) = \min(4, 20) = 4$

$D(2) = \min(|2-4|, |2-20|) = \min(2, 18) = 2$

$D(4) = \min(|4-4|, |4-20|) = \min(0, 16) = 0$

$D(18) = \min(|18-4|, |18-20|) = \min(14, 2) = 2$

$D(20) = \min(|20-4|, |20-20|) = \min(16, 0) = 0$

$\therefore C_3 = 0 \quad (k=3)$

Now that $k=3$, $C = \{C_1, C_2, C_3\} = \{0, 4, 20\}$