3.1 Problem 1 : SVM [6pts]

The "maximum margin classifier" (also called linear "hard margin" SVM) is a classifier that leaves the largest possible margin on either side of the decision boundary. The samples lying on the margin are called support vectors.

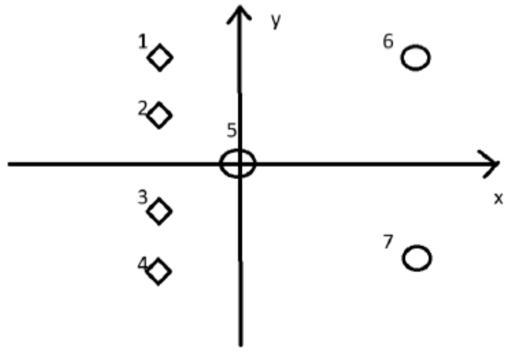
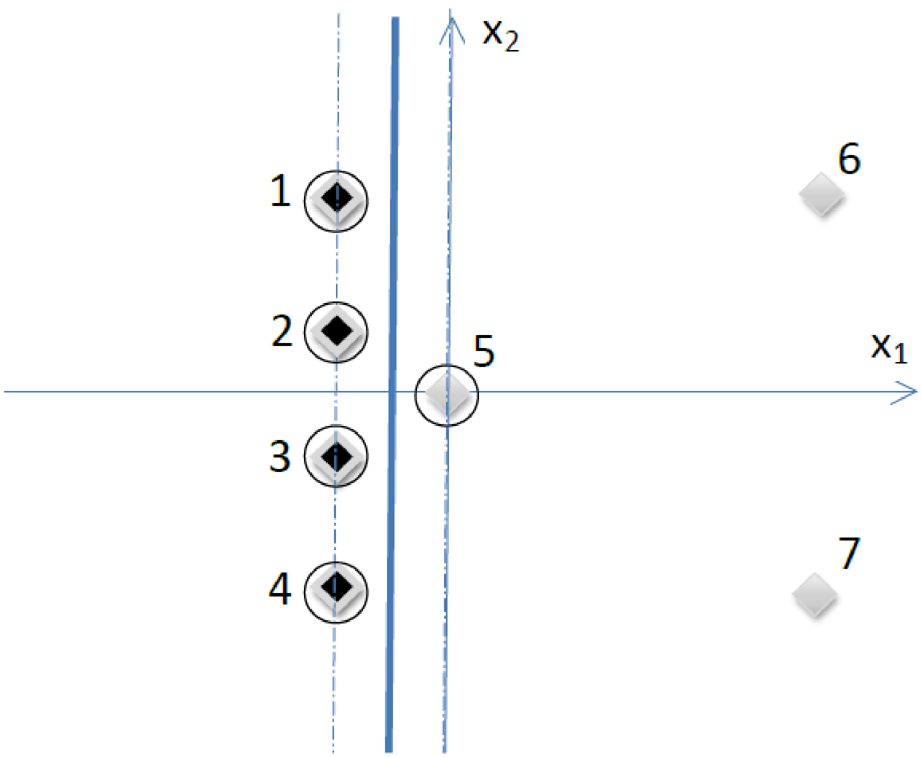


Figure 1

3.1.a Draw on Figure 1 the decision boundary obtained by the linear hard margin SVM method with a thick solid line. Draw the margins on either side with thinner dashed lines. Circle the support vectors.

3.2.b The removal of which sample will change the decision boundary?

A.



B. 5

3.2 Problem 2: k-means cluster [6pts]

Cluster the following eight points (with (x, y) representing locations) into three clusters:
A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9)

Initial cluster centers are: A1(2, 10), A4(5, 8) and A7(1, 2).

The distance function between two points a = (x1, y1) and b = (x2, y2) is defined as-
P(a, b) = |x2 – x1| + |y2 – y1|

Use K-Means Algorithm to find the three cluster centers after the second iteration.

1st iteration:

$P(A_2, A_1) = 5$ $\quad$ $P(A_3, A_1) = 12$ $\quad$ $P(A_5, A_1) = 10$

$P(A_2, A_4) = 6$ $\quad$ $P(A_3, A_4) = 7$ ✓ $\quad$ $P(A_5, A_4) = 5$ ✓

$P(A_2, A_7) = 4$ ✓ $\quad$ $P(A_3, A_7) = 9$ $\quad$ $P(A_5, A_7) = 9$

$P(A_6, A_1) = 10$ $\quad$ $P(A_8, A_1) = 3$

$P(A_6, A_4) = 5$ ✓ $\quad$ $P(A_8, A_4) = 2$ ✓

$P(A_6, A_7) = 7$ $\quad$ $P(A_8, A_7) = 10$

Cluster 1: $A_1 \Rightarrow C_1(2, 10)$

Cluster 2: $A_3, A_4, A_5, A_6, A_8 \Rightarrow C_2(6, 6)$

Cluster 3: $A_2, A_7 \Rightarrow C_3(\frac{3}{2}, \frac{7}{2})$

2nd iteration

	$C_1(2,10)$	$C_2(6,6)$	$C_3(\frac{3}{2}, \frac{7}{2})$ (1.5, 3.5)
$A_1(2,10)$	0 ✓		
$A_2(2,5)$	5	5	2 ✓
$A_3(8,4)$	12	4 ✓	7
$A_4(5,8)$	5	3 ✓	8
$A_5(7,5)$	10	2 ✓	7
$A_6(6,4)$	10	2 ✓	5
$A_7(1,2)$	9	9	2 ✓
$A_8(4,9)$	3 ✓	5	8

$\Rightarrow C_1 = (3, \frac{19}{2}),\ C_2 = (\frac{13}{2}, \frac{21}{4}),\ C_3 = (\frac{3}{2}, \frac{7}{2})$

3.3 Problem 3: PCA algorithm [6pts]

Consider the two dimensional patterns (2, 1), (3, 5), (4, 3), (5, 6), (6, 7), (7, 8).

Compute the principal component using PCA Algorithm.

① Normilize X:

$$\mu = \left(\frac{9}{2}, 5\right) = (4.5, 5) \Rightarrow X' = \begin{bmatrix} -2.5 & -4 \\ -1.5 & 0 \\ -0.5 & -2 \\ 0.5 & 1 \\ 1.5 & 2 \\ 2.5 & 3 \end{bmatrix}$$

② Covariance matrix

$$\Sigma = \frac{1}{m} X'^{T} \cdot X' = \frac{1}{6} \begin{bmatrix} -2.5 & -1.5 & -0.5 & 0.5 & 1.5 & 2.5 \\ -4 & 0 & -2 & 1 & 2 & 3 \end{bmatrix} \cdot \begin{bmatrix} -2.5 & -4 \\ -1.5 & 0 \\ -0.5 & -2 \\ 0.5 & 1 \\ 1.5 & 2 \\ 2.5 & 3 \end{bmatrix}$$

$$= \frac{1}{6} \begin{bmatrix} 17.5 & 22 \\ 22 & 34 \end{bmatrix}$$

③ eigenvalues & eigenvectors. $A\vec{x} = \lambda\vec{x}$

$$\det(\Sigma - \lambda I) = 0$$

$$\begin{vmatrix} 17.5/6 - \lambda & 22/6 \\ 22/6 & 34/6 - \lambda \end{vmatrix} = \left(\frac{17.5}{6} - \lambda\right)\left(\frac{34}{6} - \lambda\right) - \left(\frac{22}{6}\right)^2 = 0$$

$$\lambda^2 - \frac{51.5}{6}\lambda + \frac{111}{36} = 0$$

$$\Delta = \frac{8833}{144}$$

$$\lambda = \frac{\frac{51.5}{6} \pm \sqrt{8833}/12}{2}$$

$\Rightarrow \lambda_1 = 8.2077$ (principal component).

$\lambda_2 = 0.3757$

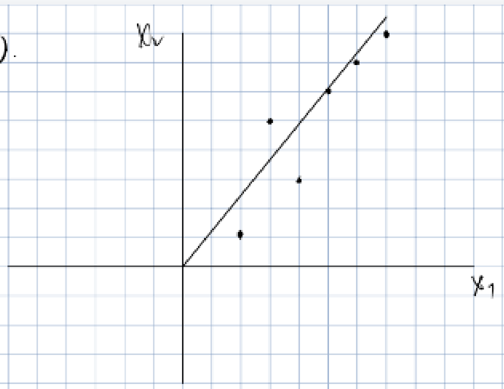
$\frac{1}{6}\begin{bmatrix} 17.5 & 22 \\ 22 & 34 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 8.2077\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$\frac{1}{6}\begin{bmatrix} 17.5x_1 + 22x_2 \\ 22x_1 + 34x_2 \end{bmatrix} = 8.2077\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$\frac{17.5}{6}x_1 + \frac{22}{6}x_2 = 8.2077x_1$

$\frac{22}{6}x_2 = 5.2910x_1$

$\Rightarrow x_1 = 0.693x_2$ eigenvector $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0.693 \\ 1 \end{bmatrix}$



3.4 Problem 4: SVM[7pts]

Assume we have 6 points showing 6 observations in a 2-D Euclidean space as below:

Observation	X_1	X_2	Y
1	0	2	class 1
2	0	3	class 1
3	1	3	class 1
4	1	1	class 2
5	2	1	class 2
6	2	2	class 2

Using a maximum margin classifier, determine the optimal separating hyperplane and give an equation for it, and determine which observations are the support vectors.

First, let's assume that we have a separating hyperplane defined as:

$$f(\vec{x}) = \beta_0 + \vec{x}^T\vec{\beta} = 0$$

where $\vec{x} = (x_1, x_2)$, $\vec{\beta} = (\beta_1, \beta_2)$.

Since we have two classes: class 1 and class 2, without loss of generality we can take $y = 1$ for class 1, and $y = -1$ for class 2.

Next let's define the sign function:

$$sign(x) = \begin{cases} 1 & x > 0 \\ -1 & x < 0 \end{cases} \tag{1}$$

Recall from lecture notes that when we fit a maximal margin classifier, with the classification rule as $G(\vec{x}) = sign(\beta_0 + \vec{x}^T\vec{\beta})$, we want to find $\beta_0, \vec{\beta}$ such that with $|\vec{\beta}| = 1$, $y_i(\beta_0 + x_{i,1}\beta_1 + x_{i,2}\beta_2) \geq M$ is true for all integer $1 \leq i \leq 6$ (for all the sample points in our toy dataset), the value of M is maximized.

Since we assume that a separating hyperplane exists, we have $M > 0$, then we have:

$$y_i(\beta_0 + x_{i,1}\beta_1 + x_{i,2}\beta_2) \geq M$$

$$y_i(\frac{\beta_0}{M} + \frac{x_{i,1}\beta_1}{M} + \frac{x_{i,2}\beta_2}{M}) \geq \frac{M}{M}$$

$$y_i(\frac{\beta_0}{M} + \frac{x_{i,1}\beta_1}{M} + \frac{x_{i,2}\beta_2}{M}) \geq 1$$

Now let's take: $\frac{\beta_0}{M} = \beta_0'$,$\frac{\beta_1}{M} = \beta_1'$, $\frac{\beta_2}{M} = \beta_2'$, and the above expression can be changed into:

$$y_i(\beta_0' + x_{i,1}\beta_1' + x_{i,2}\beta_2') \geq 1$$

Since $|\vec{\beta}| = \sqrt{\beta_1^2 + \beta_2^2} = 1$, $\sqrt{{\beta_1'}^2 + {\beta_2'}^2} = \frac{1}{M}$, maximizing M is equivalent with minimizing $\sqrt{{\beta_1'}^2 + {\beta_2'}^2}$.

Now, we can transform our optimization problem into the below linear programming styled problem:

With the constraints $y_i(\beta_0' + x_{i,1}\beta_1' + x_{i,2}\beta_2') \geq 1$ satisfied for all integer $1 \leq i \leq 6$ (for all the sample points in our toy dataset), find that value of $\beta_0', \beta_1', \beta_2'$ such that the value of $\sqrt{{\beta_1'}^2 + {\beta_2'}^2}$ is maximized.

Next we list all of our constraints:

$$\begin{aligned}
&(\beta_0' + 0\beta_1' + 2\beta_2') \geq 1 &&[1]\\
&(\beta_0' + 0\beta_1' + 3\beta_2') \geq 1 &&[2]\\
&(\beta_0' + 1\beta_1' + 3\beta_2') \geq 1 &&[3]\\
&-(\beta_0' + 1\beta_1' + 1\beta_2') \geq 1 &&[4]\\
&-(\beta_0' + 2\beta_1' + 1\beta_2') \geq 1 &&[5]\\
&-(\beta_0' + 2\beta_1' + 2\beta_2') \geq 1 &&[6]
\end{aligned}$$

[1] + [6] gives us: $\beta_1' \leq -1$

[3] + [4] gives us: $\beta_2' \geq 1$

Since we want to minimize $\sqrt{{\beta_1'}^2 + {\beta_2'}^2}$, we take $\beta_1' = -1$ and $\beta_2' = 1$.

Then, [3] becomes $\beta_0' + (-1) + 3 \times 1 \geq 1$, which is $\beta_0' \geq -1$;

[4] becomes $-(\beta_0' + (-1) + 1) \geq 1$, which is $\beta_0' \leq -1$.

Now we know that if we take $\beta_1' = -1$ and $\beta_2' = 1$, then $\beta_0' = -1$.

The last step is to verify that $\beta_0' = -1$,$\beta_1' = -1$, and $\beta_2' = 1$ satisfies all the constraints listed above:

$$\begin{aligned}
&-1 + 0 + 2 \geq 1 &&[1](satisfied, equality\ reached)\\
&-1 + 0 + 3 \geq 1 &&[2](satisfied, equality\ not\ reached)\\
&-1 - 1 + 3 \geq 1 &&[3](satisfied, equality\ reached)\\
&-(-1 - 1 + 1) \geq 1 &&[4](satisfied, equality\ reached)\\
&-(-1 - 2 + 1) \geq 1 &&[5](satisfied, equality\ not\ reached)\\
&-(-1 - 2 + 2) \geq 1 &&[6](satisfied, equality\ reached)
\end{aligned}$$

So, we get our optimal solution $\beta_0' = -1$, $\beta_1' = -1$, and $\beta_2' = 1$, which means that our best separating hyperplane has equation: $-1 - x_1 + x_2 = 0$, or $x_2 = x_1 + 1$, our classification rule is that $y_j = sign(-1 - x_{j,1} + x_{j,2})$ for some sample with index j, with $y = 1$ for class 1, and $y = -1$ for class 2, and from the inequality satisfiability check we know that observation 2 and 5 are not support vectors, while observation 1,3,4, and 6 are support vectors.