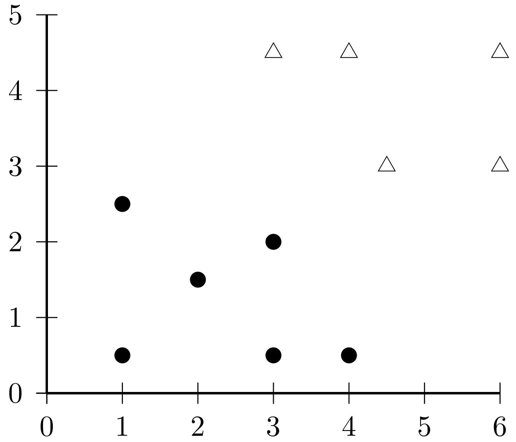
Assignment 5 - Solutions

Exercise 1 (IIR 15) [4 P.]

As you know from class, a Support Vector Machine (SVM) is estimated by finding the smallest vector $\vec{w}$ so that

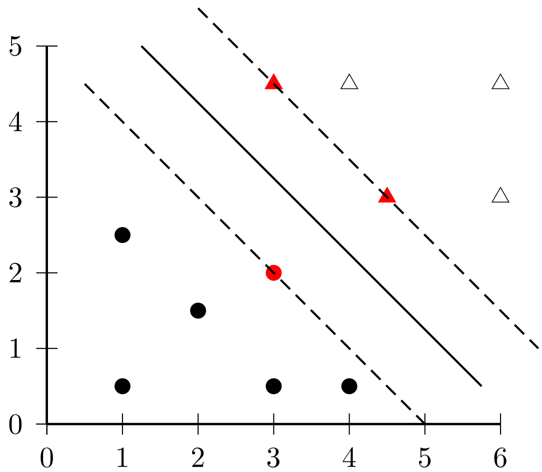
$$y_i(\vec{w}^T x_i + b) \geq 1. \quad (1)$$

Estimate a SVM for the data given below. Find the support vectors and the general form of the normal vector before solving the equation system 1 for the best matching $\vec{w}$. The Chapter 15.1 of the IR book may help you with this exercise.



Solution

(1) Correct solution:



The weight vector is parallel to the shortest line connecting points of the two classes. That is, the line between $\vec{x}_1 = (3, 2)$ and $\vec{x}_2 = (4.5, 3)$, giving a weight vector of $\vec{x}_2 - \vec{x}_1 = (1.5, 1)$. But if we take the two points $\vec{x}_1$ and $\vec{x}_2$ as support vectors for the classifier margin we see that the point $\vec{x}_3 = (3, 4.5)$ is inside the margin (see second-best solution below). Thus, the decision hyperplane constructed only from the two points $\vec{x}_1$ and $\vec{x}_2$ would not guarantee the largest margin possible. Therefore we have to consider the point $\vec{x}_3$ as an additional support vector.

Before we can arrange an equation system to find the best matching $\vec{w}$ we have to calculate the weight vector $\vec{w}$ which is perpendicular to the decision hyperplane. Since there are two support vectors from the triangles' class ($\vec{x}_2$ and $\vec{x}_3$) on the top margin line we first compute the vector that connects them: $\vec{x}_2 - \vec{x}_3 = (1.5, -1.5)^T$. Next, we compute a possible perpendicular vector by means of the dot product (we know that two vectors are perpendicular if their dot product is 0): $1.5 \cdot 1 + (-1.5) \cdot 1 = 0$. Thus, $\vec{w} = (1,1)^T$. So we already know that the solution is $\vec{w} = (1a, 1a)$ for some a. Using this knowledge we can arrange a system of equations which considers all three support vectors:

$$\begin{array}{rcrcr} (\vec{w} \cdot \vec{x}_1 + b) & = & 5a + b & = & -1 \\ (\vec{w} \cdot \vec{x}_2 + b) & = & 7.5a + b & = & 1 \\ (\vec{w} \cdot \vec{x}_3 + b) & = & 7.5a + b & = & 1 \end{array}$$

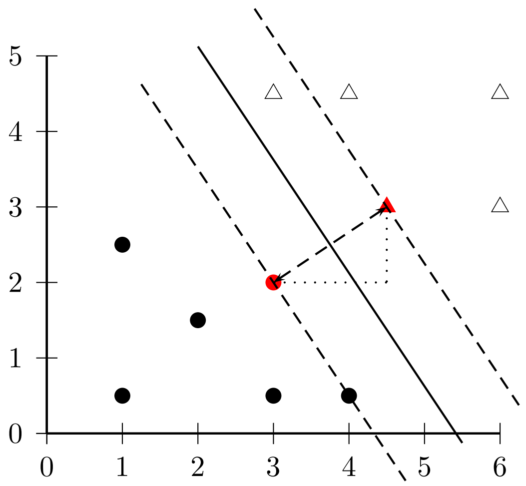
By solving this system of equations we get the following values for a und b:

$$a = \frac{4}{5}$$
$$b = -5$$

So the optimal hyperplane is given by $\vec{w} = (1 \cdot 4/5, 1 \cdot 4/5) = (4/5, 4/5)$ and $b = -5$.

Optional: The margin ρ is $2/|\vec{w}| = 2/\sqrt{0.8^2 + 0.8^2} \approx 2/\sqrt{1.28} \approx 2/1.131 \approx 1.77$.

(2) Second-best solution:



The weight vector is parallel to the shortest line connecting points of the two classes. That is, the line between $\vec{x}_1 = (3,2)$ and $\vec{x}_2 = (4.5,3)$, giving a weight vector of $\vec{x}_2 - \vec{x}_1 = (1.5, 1)$. In this (not completely correct) solution we ignore that there is another point of the triangles' class that is within the functional margin.

Thus, we know that $\vec{w} = (1.5a, 1a)$ for some a. So we get the following system of equations:

$$\begin{array}{rcrcr} (\vec{w} \cdot \vec{x}_1 + b) & = & -6.5a + b & = & -1 \\ (\vec{w} \cdot \vec{x}_2 + b) & = & 9.75a + b & = & 1 \end{array}$$

By solving the system of equations we get the following values for a und b:

$$a = \frac{8}{13}$$
$$b = -5$$

So the optimal hyperplane is given by $\vec{w} = (1.5 \cdot 8/13, 1 \cdot 8/13) = (12/13, 8/13) \approx (0.923, 0.615)$ and $b = -5$.

Optional: The margin ρ is $2/|\vec{w}| = 2/\sqrt{0.923^2 + 0.615^2} \approx 2/\sqrt{1.23} \approx 2/1.109 \approx 1.803$.

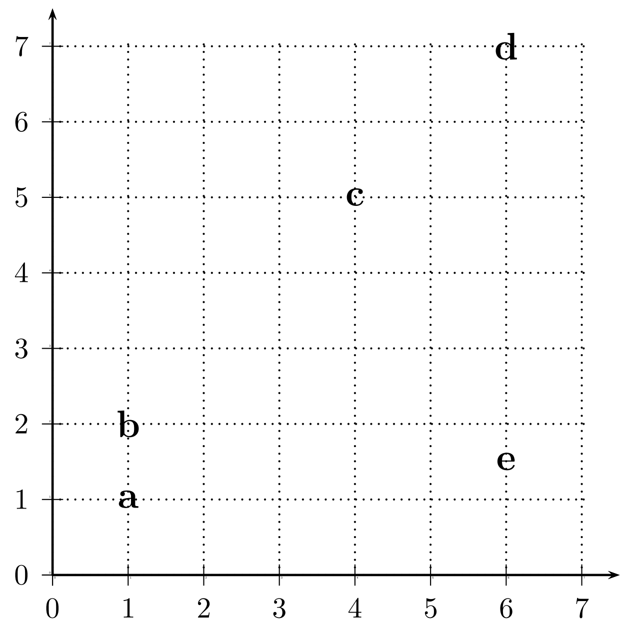
Exercise 2 (IIR 16) [4 P.]

(i) Perform a 2-means clustering to convergence for the points below. Start with the two seeds a and b. For each iteration give (I) the coordinates of the centroids (II) the assignments of points to centroids.
(ii) Give the coordinates of a fifth point e and two centroids with the following properties:

1. the two centroids are a local optimum; that is, one iteration of reassignment and recomputation will not change the position of the centroids
2. the two centroids are not the global optimum.

(iii) Give two centroids that are better for the 5 points than the ones in (ii). (No need to prove global optimality, but show they are better than in (ii).)

Solution



(i)
1. centroids A = (1,1), B = (1,2)
1. assignment: A: a; B: b, c, d
2. centroids A = (1,1) B = (11/3,14/3) = (3.67,4.67)
2. assignment: A: a,b; B: c,d
3. centroids: A: (1,1.5) B: (5,6)
3. assignment A: a,b; B: c,d
4. centroids: A: (1,1.5) B: (5,6)
$\Longrightarrow$ converged

(ii)
$e = (6, 1.5)$, A = (3, 2.375), D = (6, 7)
$|A, c|^2 = (3-4)^2 + (2.375-5)^2 \approx 7.89$
$|D, c|^2 = (6-4)^2 + (7-5)^2 = 8$
So c is closer to A than to D. a, b and e are also closer to A than to D. So this is a stable set of two centroids for the five points.
RSS = sum of squared distances of a, b, c, d, e (where a, b, c, e are assigned to A and d is assigned to D): $((3-1)^2+(2.375-1)^2)+((3-1)^2+(2.375-2)^2)+7.89+(0^2+0^2)+((3-6)^2+(2.375-1.5)^2) \approx 27.69$

(iii)
A = (8/3, 1.5), D = (5, 6) is a better set of centroids than the ones in (ii). Proof:
RSS = sum of squared distances of a, b, c, d, e (where a, b, e are assigned to A and c, d are assigned to D): $((8/3-1)^2+0.5^2)+((8/3-1)^2+0.5^2)+(1^2+1^2)+(1^2+1^2)+((8/3-6)^2+0^2) \approx 21.17 < 27.69$

Exercise 3 (IIR 16) [1 P.]

Why are documents that do not use the same term for the concept *car* likely to end up in the same cluster in *K-means* clustering?

Solution

If two document are thematically similar they contain similar terms. Even if a frequent term does not occur in a document the document will be correctly clustered due to other terms common to the topic.

Exercise 4 (IIR 16) [1 P.]

Two of the possible termination conditions for K-means were (1) assignment does not change, (2) centroids do not change (IR book, page 361). Do these two conditions imply each other? Why or why not?

Solution

The two conditions imply each other. If the assignment does not change then the centroids remain the same. And if the centroids do not change that means that no reassignment took place.