Solutions - MFML/MFDS Comprehensive Makeup

Q1 Answer the following questions with justifications.

(A) A signal tower is placed at $(0,0)$. The cost to place a receiver at a point (x,y) is equal to the square of the Euclidean distance from that point to the point at which the signal tower is placed. The region of interest is $\{(x,y)|7-2x+y \leq 0, 0 \leq x+y+5\}$.

i) Help the analyst to formulate the optimization problem to find the point at which the cost to place the receiver is minimum in the region of interest. (1.5Marks)

ii) Write down the Lagrangian for the constrained optimization problem so obtained in i) (0.5 Marks)

iii) What are the KKT conditions for the problem in i). (4 Marks)

iv) Using KKT conditions, find the solution for the optimization problem in i). (4 Marks)

Solution The formulation is

$$\begin{aligned} & \min x^2+y^2 \\ \text{subject to constraint} \quad & 7-2x+y \leq 0, \\ & x+y+5 \geq 0 \end{aligned}$$

(0.5 Marks for objective function, 1 Mark for constraint.)
The Lagrangian is given by

$$L(x,y,\lambda_1,\lambda_2) = x^2+y^2+\lambda_1(7-2x+y)+\lambda_2(-5-x-y).$$

(0.5 Marks)
If the optimum values are obtained at $x, y, \lambda_1, \lambda_2$ then the KKT conditions are

$$\begin{aligned} 7-2x+y &\leq 0, \text{ (0.5 Marks)} \\ -x-y-5 &\leq 0, \text{ (0.5 Marks)} \\ \lambda_1, \lambda_2 &\geq 0, \text{ (1 Mark)} \\ \lambda_1(7-2x+y) &= 0 \text{ (0.5 Marks)} \quad (1) \\ \lambda_2(-5-x-y) &= 0 \text{ (0.5 Marks)} \quad (2) \\ \frac{\partial L}{\partial x} &= 2x-2\lambda_1-\lambda_2 = 0 \\ \Rightarrow x &= \frac{2\lambda_1+\lambda_2}{2} \text{ (0.5 Marks)} \\ \frac{\partial L}{\partial y} &= 2y+\lambda_1-\lambda_2 = 0 \\ \Rightarrow y &= \frac{-\lambda_1+\lambda_2}{2} \text{ (0.5 Marks)} \end{aligned}$$

Substituting the values of x and y in the equations (1) and (2) we get

$$\begin{aligned} \lambda_1(7 - 2.5\lambda_1 - 0.5\lambda_2) &= 0 \\ \lambda_2(5 + 0.5\lambda_1 + \lambda_2) &= 0 \end{aligned}$$

(1 Mark)
Solving for λ_i, we get $(\lambda_1, \lambda_2) = (0,0), (2.8,0), (0,-5), (4.2,-7.1)$ (1 Mark).
$(0,0)$ will give $(x,y) = (0,0)$ which does not satisfy the 1st constraint. Therefore, $(\lambda_1, \lambda_2) = (2.8, 0) \Rightarrow (x,y) = (2.8, -1.4)$ gives the optimal solution. (2 Marks).

Q Answer the following questions with justifications.

A Consider the function $f(x) = ax^3 + bx^2 + cx + d$ where $a > 0$.

(a) Find the condition on a, b, c such that the function has two distinct critical points. Calculate the critical points in terms of a, b, c. Identify the nature of each critical point (i.e maxima or minima) [3 Marks]

(b) Define a zone of attraction around each local minimum to be the region around it such that if gradient descent starts at any point in the region, it would end up at the given local minimum. Find the zone of attraction for each local minimum, if any, of the critical points. Justify your answer mathematically. [3 Marks]

Solution:

For part (a): To find the critical points we take $\frac{df}{dx} = 3ax^2 + 2bx + c = 0$, to obtain two roots $x_1 = \frac{-b-\sqrt{b^2-3ac}}{3a}$ and $x_2 = \frac{-b+\sqrt{b^2-3ac}}{3a}$. In order for the the two roots to be real and distinct, the quantity under the square-root sign needs to be strictly positive, i.e $b^2 - 3ac > 0$. We see that the second derivative $\frac{d^2f}{dx^2} = 6ax + 2b$. The second derivative is negative for the critical point x_1 and positive for x_2 which means x_1 is a maxima and x_2 is a minima.

[3 Marks]

For part (b): As identified in part (a), there is a single local minimum $x_2 = \frac{-b+\sqrt{b^2-3ac}}{3a}$. We can rewrite $\frac{df}{dx} = 3a(x + \frac{b}{3a})^2 + (c - \frac{b^2}{3a})$. Solving for $\frac{df}{dx} < 0$ we see that $\frac{df}{dx} < 0$ when $x > \frac{-b-\sqrt{b^2-3ac}}{3a}$ and $x < \frac{-b+\sqrt{b^2-3ac}}{3a}$. When $x > \frac{-b+\sqrt{b^2-3ac}}{3a}$, $\frac{df}{dx} > 0$. Since the derivative is negative on side of the local minimum and positive on the other side of it, we see that gradient descent will take any point on the left of the local minimum x_2 but greater than $\frac{-b-\sqrt{b^2-3ac}}{3a}$ to x_2 in a sufficiently large number of steps with a suitable step-size. Similarly since the derivative is positive on the right of x_2, gradient descent will take any point on the right of x_2 to x_2 in a sufficiently large number of steps with a suitable step-size. Thus the zone of attraction for the local minimum $x_2 = \frac{-b+\sqrt{b^2-3ac}}{3a}$ is $\frac{-b-\sqrt{b^2-3ac}}{3a} < x \leq \infty$.

[3 Marks]

B Consider solving the following constrained optimization problem:

$$\min ax + by + cz \text{ subject to } px + qy + rz = 0$$

Use the theory of Lagrange multipliers to obtain a condition on (a, b, c) and (p, q, r) when there is a solution to this problem and when there is no solution. How many solutions are there to this problem when there is a solution, and what is the value of the objective function at a solution point? (4 Marks)

Solution:

We first set up the Lagrangian $L(x, y, z, \lambda) = ax + by + cz - \lambda(px + qy + rz)$. Setting $\frac{\partial L}{\partial x} = 0$, $\frac{\partial L}{\partial y} = 0$, $\frac{\partial L}{\partial z} = 0$, we get $a = \lambda p$, $b = \lambda q$ and $c = \lambda r$. For there to be a value of λ that solves all these three equations, we need to have $\frac{a}{p} = \frac{b}{q} = \frac{c}{r}$. This means that for their to be a critical point (a, b, c) and (p, q, r) point in the same or opposite directions. On the other hand when (a, b, c) and (p, q, r) point in different directions, there is no solution to λ and there is no critical point which means there is no point at which the minimum is attained. When there exists a solution, any point on the surface $px+qy+rz = 0$ is a solution and the value of the objective function at a solution point (x_0, y_0, z_0) is $k(ax_0+by_0+cz_0) = 0$.

[4 Marks]

Q Answer the following with justifications:

(A) You are given a design matrix:

$$\mathbf{X} = \begin{bmatrix} 6 & -4 \\ -3 & 5 \\ -2 & 6 \\ 7 & -3 \end{bmatrix}$$

(a) Compute the covariance matrix for the sample points. Then compute the unit eigenvectors, and the corresponding eigenvalues, of the covariance matrix.

(4 marks)

(b) Use PCA to project the sample points onto a one-dimensional space. What one-dimensional subspace are we projecting onto? For each of the four sample points in X, write the coordinate (in principal coordinate space, not in R^2) that the point is projected to.

(3 marks)

(c) Draw the rough sketch of the projected points on the principal axis.

(1 mark)

Solution

(A)

x	y	(x - 2)	(y - 0.5)	(x-2) ^2	(y -0.5) ^2	(x-2) (y-0.5)
6	-4	4	-4.5	16	20.25	-18
-3	5	-5	4.5	25	20.25	-22.5
-2	6	-4	5.5	16	30.25	-22
7	-5	5	-5.5	25	30.25	-27.5
Mean = 2	**Mean = 0.5**			**82**	**101**	**-90**

Covariance Matrix:

$$\Sigma = \frac{1}{N}\begin{bmatrix} Cov\ (x,x) & Cov\ (x,y) \\ Cov\ (y,x) & Cov\ (y,y) \end{bmatrix} = \frac{1}{4}\begin{bmatrix} 82 & -90 \\ -90 & 101 \end{bmatrix} = \begin{bmatrix} 20.5 & -22.5 \\ -22.5 & 25.25 \end{bmatrix}$$ 1.5

Eigen values: $|\Sigma - \lambda I| = 0 \Rightarrow \lambda_1 = 45.5$, $\lambda_2 = 0.25$ 0.5

Eigen vectors: $v_1 = \begin{bmatrix} -0.9 \\ 1 \end{bmatrix}$ and $v_2 = \begin{bmatrix} 1.11 \\ 1 \end{bmatrix}$ 2

Unit Eigenvectors: $\hat{e}_1 = \frac{1}{1.35}\begin{bmatrix} -0.9 \\ 1 \end{bmatrix} = \begin{bmatrix} -0.67 \\ 0.74 \end{bmatrix}$ and $\hat{e}_2 = \frac{1}{1.49}\begin{bmatrix} 1.11 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.74 \\ 0.67 \end{bmatrix}$

(B)

The one-dimensional subspace we are projecting onto is along the **principal eigenvector**: 1

$\hat{e}_1 = \begin{bmatrix} -0.67 \\ 0.74 \end{bmatrix}$ which corresponds to the direction of maximum variance in the data.

The coordinates (in principal coordinate space) for each of the four sample points are projected as $Y = \hat{e}_1{}^T X$ and are given by: 2

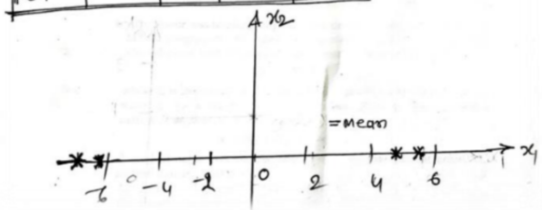
$(6, -4) \rightarrow 6.99$, $(-3, 5) \rightarrow -5.72$, $(-2, 6) \rightarrow -5.80$, $(7, -3) \rightarrow 6.91$

(C)

Geometry

$PC_1 = e_1^T x = -0.67 x_1 + 0.74 x_2$

x_1	6	-3	-2	7
x_2	-4	5	6	-3
PC1	-6.98	5.71	5.78	-6.91

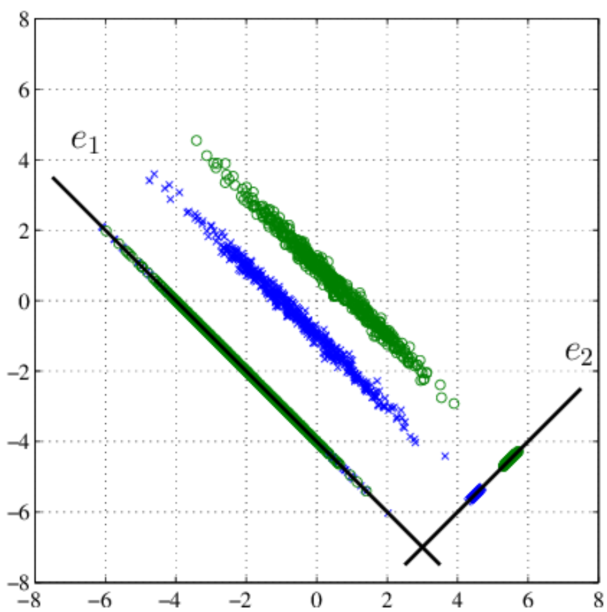


1

(B) Assume that you have data points belonging to two classes. Will it always be true that if the data is projected on the first principal component, it can be classified into two classes by a simple thresholding? Explain why it is true, or provide a counter example if it is false.

(2 mark)

Solution No. counter example:



Q Answer the following with justifications:

(A) Assume that you received the below data from a noisy sensor. The data consists of two classes (* and o). You train an SVM classifier with a quadratic kernel (i.e., polynomial with degree 2). Assume that C is the regularisation parameter used to build a hard margin / soft margin classifier. Draw a rough replica of the below image and highlight your answers on that image. Give proper justification for each of your answers for below questions.

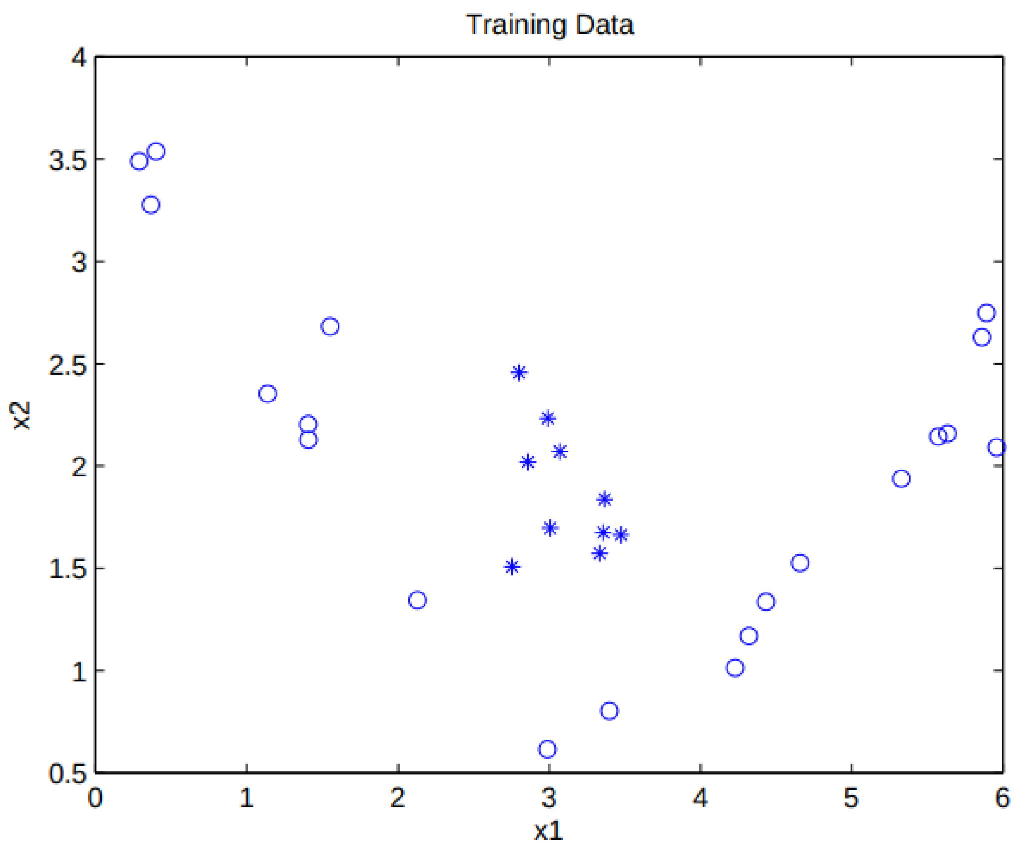


FIGURE 1. Training Data in $\mathbb{R}^2$. Number of classes = 2

(a) Draw a possible decision boundary for a small value of C?

(b) Draw a possible decision boundary for a large value of C?

(c) Highlight one point, if it had a wrong label when given to you, would *not* affect the decision boundary

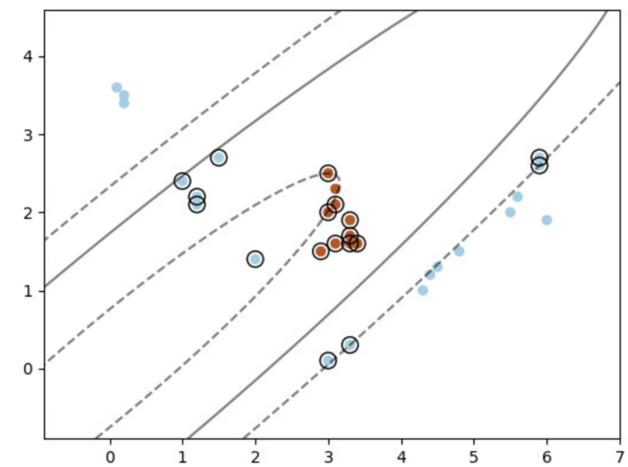
(d) Highlight one point, if it had a wrong label when given to you, would affect the decision boundary

(e) Would you prefer a high value of C or a low value of C, given the nature of the problem?

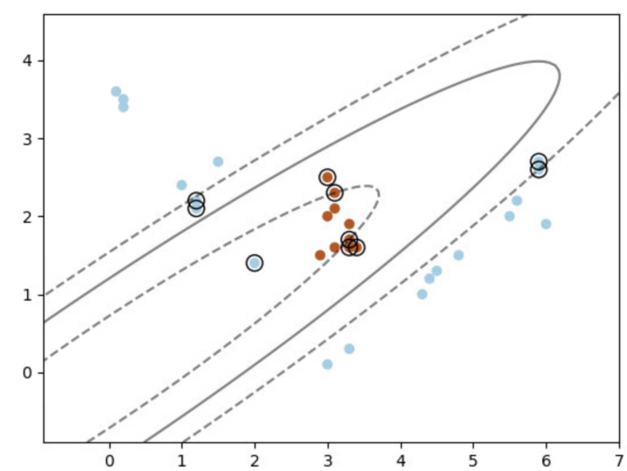
(1 mark x 5)

Solution:

a one possible decision boundary for small C



b one possible decision boundary for large C



With large C, misclassification should be less. Some support vectors should be highlighted other orientations of classification boundary is also acceptable

c one of the points NOT in the support vector to be highlighted

d one of the points in the support vector to be highlighted (award marks if the this is highlighted in either of the two plots for low C and high C)

e given that the data is from a noisy sensor, we should accommodate for some misclassification. So C should be low.

[1 mark each]

(B) A function $f : \mathbb{R}^3 \Rightarrow \mathbb{R}^3$ is given below:

$$\mathbf{f([x_1, x_2, x_3]^T) = [2 * x_1 + x_2, 3 * x_1 + 7 * x_2 + 3 * x_3, 3 * x_1 + x_2 + x_3]^T}$$

(A) Using elementary row operations, write the matrix in its row echelon form

(2 marks)

(B) Let $\mathbf{V}$ be a vector subspace spanned by the columns of matrix $\mathbf{A}$. Give a set of vectors that form the basis for the vector space $\mathbf{V}$. What is the number of vectors in this basis set?

(1 marks)

(C) Let $\mathbf{V}$ be a vector subspace spanned by vectors $\mathbf{x}$, such that $\mathbf{Ax} = 0$. Give a set of vectors that form the basis for the vector space $\mathbf{V}$. What is the number of vectors in this basis set?

(1 marks)

(D) Give the set of linearly independent rows of $\mathbf{A}$. What is the number of vectors in this set?

(1 marks)

A answer with marking schemes.

(A)

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & 0 \\ 3 & 7 & 3 \\ 3 & 1 & 1 \end{bmatrix}$$

(2 marks)

(B)

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & 0 \\ 3 & 7 & 3 \\ 3 & 1 & 1 \end{bmatrix}$$

$$\xrightarrow{R_1 \leftarrow R_1/2, R_2 \leftarrow R_2 - 3 * R_1, R_3 \leftarrow R_3 - 3 * R_1, R_2 \leftarrow R_2/(11/2), R_3 \leftarrow R_3/(-1/2), R_3 \leftarrow R_3 - R_2}$$

$$\begin{bmatrix} 0 & \frac{1}{2} & 0 \\ 0 & 1 & \frac{6}{11} \\ 0 & 0 & \frac{-17}{11} \end{bmatrix}$$

NOTE: REF is not unique. 1 mark

(C) All columns have pivot. All columns form the basis of column space

$$\begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix} \begin{bmatrix} 1 \\ 7 \\ 1 \end{bmatrix} \begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix}$$

Rank of col space = 3

(1 mark)

(C) From REF, the null space is: $\{\mathbf{0}\}$
Rank of null space = 0 (1 mark