---

(1) Find the LU decomposition of the following 3×3 matrix A with L in lower triangular matrix form and U as upper triangular form with unit diagonal elements:

$$A = \begin{pmatrix} 3 & 6 & -9 \\ 2 & 5 & -3 \\ -4 & 1 & 10 \end{pmatrix}$$

Solution:

From

$$L = \begin{pmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{pmatrix} \quad U = \begin{pmatrix} 1 & u_{12} & u_{13} \\ 0 & 1 & u_{23} \\ 0 & 0 & 1 \end{pmatrix}$$

We get 9 equations:

$$\begin{aligned} l_{11} &= 3 \\ l_{11}u_{12} &= 6 \\ l_{11}u_{13} &= -9 \\ l_{21} &= 2 \\ l_{21}u_{12} + l_{22} &= 5 \\ l_{21}u_{13} + l_{22}u_{23} &= -3 \\ l_{31} &= -4 \\ l_{31}u_{12} + l_{32} &= 1 \\ l_{31}u_{13} + l_{32}u_{23} + l_{33} &= 20 \end{aligned}$$

which gives

$$\begin{aligned} l_{11} = 3, l_{21} = 2, l_{22} = 1, l_{31} &= -4 \\ l_{32} = 9, l_{33} = -29, u_{12} = 2, u_{13} = -3, u_{23} &= 3 \end{aligned}$$

Hence,

$$L = \begin{pmatrix} 3 & 0 & 0 \\ 2 & 1 & 0 \\ -4 & 9 & -29 \end{pmatrix} \quad U = \begin{pmatrix} 1 & 2 & -3 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix}$$

Suggested Marking Scheme: 3 Marks $\rightarrow$ setting up the equations correctly, 3 Marks $\rightarrow$ final solution

(2) Consider a d dimensional data where σ_i^2 is the variance of the i^{th} feature. The eigenvectors of unit magnitude of the data covariance matrix are $\{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_d\}$ where $\boldsymbol{e}_i \in \mathbb{R}^d$ is the vector with i^{th} component as 1 and rest all 0s and corresponds to the i^{th} largest eigenvalue.

i) Compute the data covariance matrix if $\sigma_i^2 > \sigma_j^2$ when $i < j$ with proper justification.

[2 Marks]

ii) If in the above case $d = 4$ and $\sigma_i^2 = 5 - i, i = 1, 2, 3, 4,$ then help the data scientist to eliminate one of the feature such that after elimination, 90% of the total variance is retained.

[2 Marks]

iii) If for a d dimensional data, from the d principal components, projection of the data onto first $d - 1$ principal components retains 100% of the total variance, is it possible to find the inverse of the data covariance matrix? Justify your answer.

[2 Marks]

Solution **(Kindly award marks for any alternate correct method appropriately)**

i) Let $\boldsymbol{P} = [\boldsymbol{e}_1, \cdots, \boldsymbol{e}_d]$ and $\boldsymbol{D} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ 0 & 0 & \cdots & \lambda_d \end{bmatrix}$

where λ_i is the i^{th} largest eigenvalue.
Then the data covariance matrix $\boldsymbol{S} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1} = \boldsymbol{D}$ as $\boldsymbol{P}$ is an n dimensional identity matrix. [1 Mark]
Therefore, $\lambda_i = \sigma_i^2, \ \forall i = 1, \cdots, d$ and

hence $\boldsymbol{S} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ 0 & 0 & \cdots & \sigma_d^2 \end{bmatrix}$, [1 Mark]

ii) Clearly $\dfrac{\sum_{i=1}^{3} \lambda_i}{\sum_{i=1}^{4} \lambda_i} = \dfrac{\sum_{i=1}^{3} \sigma_i^2}{\sum_{i=1}^{4} \sigma_i^2} = 0.9.$ [0.5 Marks]

$\Rightarrow$ if $\boldsymbol{X}$ is the data matrix of order $4 \times n$ where n is the number of data points then $\boldsymbol{B}\boldsymbol{B}^T\boldsymbol{X}$ gives the reconstructed data where

$$\boldsymbol{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Now $\boldsymbol{B}\boldsymbol{B}^T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \Rightarrow \boldsymbol{B}\boldsymbol{B}^T\boldsymbol{X} = \begin{bmatrix} X_{11} & \cdots & X_{1n} \\ X_{21} & \cdots & X_{2n} \\ X_{31} & \cdots & X_{3n} \\ 0 & \cdots & 0 \end{bmatrix}.$

Thus, 4^{th} feature can be eliminated in order to reduce the dimension by 1 but still retaining 90% of the total variance.

[1 Mark] for the reason and [0.5 Marks] for the correct feature.

iii) If for a d dimensional data, from the d principal components, projection of the data onto first $d - 1$ principal components retains 100% of variance implies $\dfrac{\sum_{i=1}^{d-1} \lambda_i}{\sum_{i=1}^{d} \lambda_i} = 1$ where λ_i is the i^{th} largest eigenvalue of the data covariance matrix $\boldsymbol{S}$

[0.5 Marks]

$\Rightarrow \sum_{i=1}^{d-1} \lambda_i = \sum_{i=1}^{d} \lambda_i$

$\Rightarrow \lambda_d = 0$ [0.5 Marks]
$\Rightarrow \det(\boldsymbol{S}) = \prod_{i=1}^{d} \lambda_i = 0$ [0.5 Marks]
$\Rightarrow \boldsymbol{S}$ is not invertible. [0.5 Marks]

(3) Using Lagrangian multipliers, find the maximum and minimum values of the function as applicable: $f(x;y) = x^2 + y$ and ; $x^2 - y^2 = 1$. List out the corresponding points of maxima and minima. Does the method provides inferences on both maxima and minima? Justify

[8 Marks]

Solution: We have $f_x = 2x, g_x = 2x; f_y = 1, g_y = -2y$
Solving multiplier equations, we get $f_x = \lambda gx \implies 2x = \lambda(2x)$
$f_y = \lambda g_y \implies 1 = \lambda(-2y)$ subjected to the given constraint: $g(x, y) = x^2 - y^2 = 1$
$\implies \lambda = 1$ or $x = 0$
On the other hand, $\lambda = 1$, then we get $y = \frac{-1}{2} and x = \pm\frac{\sqrt{5}}{4}$
Here x cannot be 0 because it violates the constraint equation
Therefore the points are $A(\sqrt{\frac{5}{4}}, -\frac{1}{2}); B(-\sqrt{\frac{5}{4}}, -\frac{1}{2})$
In both cases the function value turns out to be $\frac{3}{4}$
Also we have $f(x, y) = (y^2 + 1) + y = y^2 + y + 1$ (substituting the constraint equation)
Hence $f(x; y) = (y + \frac{1}{2})^2 + \frac{3}{4}$
Hence these points represent minima.

Suggested Marking Scheme: Step marking as appropriate

(4) i) Consider a quadratic function $f(x, y) = 2x^2 + \beta y^2$ where $\beta \in \mathbb{R}$ is an unknown constant. Also assume that $\beta \neq 0$. Consider the problem of minimizing this function using gradient descent algorithm.

(i) Derive a closed form expression for the optimal step size α for the first iteration of gradient descent if the initial point is

$$\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

(ii) If it is known that optimal step size $\alpha = 0.25$, derive the value of constant β using the formula derived in (a).

(4 marks)

ii) Consider the polynomial function $f(z) = 3z^4 - 6z^3 - 42z^2 + 2z + 72$.

(i) Derive all the critical points of this function in the range $[a \;\; b]$ where $a = -5$ and $b = 5$.

(ii) Using second derivative test, classify the critical point closest to the left end of the range (i.e 'a') as maxima or minima.

(2 marks)

Solution

i) a) Note that

$$\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

The gradient of $f(x, y) = 2x^2 + \beta y^2$ at (x_0, y_0) is as follows

$$\nabla f(x_0, y_0) = \begin{bmatrix} 4x_0 \\ 2\beta y_0 \end{bmatrix} = \begin{bmatrix} 4 \\ 4\beta \end{bmatrix}$$

The optimal step size is obtained as solution of following 1 dimensional optimization problem

$$\text{argmin}_\alpha f\left(\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} - \alpha \nabla f(x_0, y_0) \right)$$

(0.5 mark)

after substituting $x_0 = 1$ and $y_0 = 2$, we get that, this is equivalent to solving

$$\text{argmin}_\alpha f(1 - 4\alpha, 2 - 4\alpha\beta)$$

In other words we need to find the minimum of

$$g(\alpha) = 2(1 - 6\alpha)^2 + \beta(2 - 4\alpha\beta)^2$$

(1 mark)

To find minimum of $g(\alpha)$ with respect to α, we find $g'(\alpha)$ and equate it to zero

$$g'(\alpha) = 32\beta^3\alpha - 16\beta^2 + 64\alpha - 16 = 0$$

(1 mark)

Hence the closed form expression for α is given by

$$\alpha = \frac{16 + 16\beta^2}{64 + 32\beta^3}$$

(0.5 mark)

b) It is given that $\alpha = 0.25$, substituting in the previous expression we get the equation

$$0.25 = \frac{16 + 16\beta^2}{64 + 32\beta^3}$$

Rearranging we get the expression $\beta^2(\beta - 2) = 0$. Hence potential value of β is 0 or 2. Its given in question that $\beta \neq 0$. Hence final answer is $\beta = 2$. (1 mark)

ii) (a) The derivative is given by $f'(z) = 12z^3 - 18z^2 - 84z$. Equating it to zero we get $12z^3 - 18z^2 - 84z = 0$. This can be factorized as $6z(2z^2 - 3z - 14) = 0$

So $z = 0$ or $12z^2 - 18z - 84 = 0$. The second equation is quadratic in nature. The solution can be btained by using the formula for solving quadratic equations:

$$z = \frac{3 \pm \sqrt{(-3)^2 - 4 * 2 * (-14)}}{4}$$

The three critical points are $z_1 = 0, z_2 = -2, z_3 = 3.5$

(b) The critical point closest to a is $z_2 = -2$. The second derivative at z_2 is $f''(z) = 36z^2 - 36z - 84 = 36(-2)^2 - 36(-2) - 84 > 0$. Hence its a point of minimum.

(5) We are given a primal optimization problem D of the form

$$\begin{aligned} \min\, x + y \text{ subject to} \\ \alpha x^2 + \beta y^2 \leq 1 \\ \beta x^2 + \alpha y^2 \leq 1 \end{aligned}$$

It is known that $\alpha > 0, \beta > 0$, $\alpha \neq \beta$, and $f(x)$ is convex.

(a) Write the dual formulation for this problem. The objective function in the dual formulation should be expressed in the form $a(\lambda) + \frac{1}{p(\lambda)} + \frac{1}{q(\lambda)}$ where $a(\lambda), p(\lambda), q(\lambda)$ are all linear functions. (4 Marks)
.

(b) Assuming that the dual optimal solution is strictly positive find the optimal solution to the primal optimization problem D. (4 Marks)

Solution

(a) The Lagrangian for the problem is $x+y+\lambda_1(\alpha x^2+\beta y^2-1)+\lambda_2(\beta x^2+\alpha y^2-1)$. To calculate the objective function of the dual we need to compute $D(\lambda) = \min_{x,y} L(x, y, \lambda)$. Thus we set $\frac{\partial L}{\partial x} = 0, \frac{\partial L}{\partial y} = 0$ and obtain the equations

$$\begin{aligned} 1 + 2x(\alpha\lambda_1 + \beta\lambda_2) = 0 \\ 1 + 2y(\beta\lambda_1 + \alpha\lambda_2) = 0 \end{aligned}$$

Solving for x, y in the above equations and plugging these values into $L(x, y, \lambda)$ and simplifying we obtain $D(\lambda) = -\lambda_1 - \lambda_2 - \frac{1}{4(\alpha\lambda_1+\beta\lambda_2)} - \frac{1}{4(\alpha\lambda_2+\beta\lambda_1)}$ which is of the required form. Thus the dual formulation is

$$\begin{aligned} \max D(\lambda) \text{ subject to} \\ \lambda \geq 0 \end{aligned}$$

Suggested Marking Scheme: 2 Marks $\rightarrow$ setting up the Lagrangian correctly, 2 Marks $\rightarrow$ rest of the argument.

(b) The given problem is convex since the objective function is convex and the two given constraints are also convex due to α and β both being greater than zero. Also there exists a point interior to the primal feasible region. Therefore strong duality holds and the KKT conditions can be set up for this problem. We use complementary slackness conditions and obtain

$$\begin{aligned} \lambda_1(\alpha x^2 + \beta y^2 - 1) = 0 \\ \lambda_2(\beta x^2 + \alpha y^2 - 1) = 0 \end{aligned}$$

Since we know that the dual optimal solution is strictly greater than zero it means

$$\begin{aligned} \alpha x^2 + \beta y^2 = 1 \\ \beta x^2 + \alpha y^2 = 1 \end{aligned}$$

Subtracting the two equations gives $(\alpha - \beta)x^2 = (\alpha - \beta)y^2$. Since it is given that $\alpha \neq \beta$, we must have $x = \pm y$. Then we get $x^2 = \frac{1}{\alpha+\beta}$ which gives the minimum value of the objective function of D as

$\frac{-2}{\sqrt{\alpha+\beta}}$.

Suggested Marking Scheme: 2 Marks $\rightarrow$ recognizing that the problem is convex and strong duality holds. 2 Marks $\rightarrow$ using complementary slackness to complete the argument.

(6) Let $K(x,y)$ be a kernel function corresponding to the mapping ϕ which has d components. Find the mapping for the kernel function $(K(x,y)+c)^2$? How many components does this mapping contain? [6 Marks]

Solution

Expanding $(k(x,y)+c)^2$, we get $k^2(x,y)+2ck(x,y)+c^2)$. We can view the new transformation ψ to be composed of the concatenation of three transformations such that $\psi(x) = \begin{bmatrix}\alpha_1(x)\\ \alpha_2(x)\\ \alpha_3(x)\end{bmatrix}$, where $\alpha_1(x)$ corresponds to the kernel c^2, $\alpha_2(x)$ corresponds to the kernel $2ck(x,y)$ and $\alpha_3(x)$ corresponds to $k^2(x,y)$.

We can see that the transformation $\alpha_1(x) = c$,

$$\alpha_2(x) = [\sqrt{(2c)}\phi_1(x), \sqrt{(2c)}\phi_2(x), \ldots \sqrt{(2c)}\phi_d(x)]$$

and

$$\alpha_3(x) = [\phi_1(x)\phi_1(x), \phi_1(x)\phi_2(x)\ldots\phi_1(x)\phi_d(x), \phi_2(x)\phi_1(x), \phi_2(x)\phi_2(x), \\ \ldots\phi_2(x)\phi_d(x), \ldots\phi_d(x)\phi_1(x), \phi_d(x)\phi_2(x), \ldots\phi_d(x)\phi_d(x)]^T$$

The total number of components in the transformation ψ is $1+d+d^2$

Suggested Marking Scheme: 3 Marks $\rightarrow$ recognizing that the required transformation is composed of the concatenation of three transformations, 3 Marks $\rightarrow$ rest of the argument. .