LAHORE UNIVERSITY OF MANAGEMENT SCIENCES
Department of Electrical Engineering

AI 501 Mathematics for Artificial Intelligence
Quiz 06 Solutions

Name: ____________________

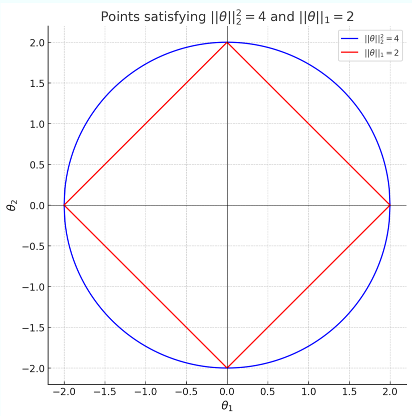
Campus ID: ____________________

Total Marks: 10
Time Duration: 15 minutes

Question 1 (2 marks)

We use the Euclidean and Manhattan norms to define the regularization term in Ridge and Lasso regression techniques respectively. Consider $\boldsymbol{\theta} \in \mathbb{R}^2$. Plot all the points in (θ_1, θ_2) plane for which $||\boldsymbol{\theta}||_2 = 2$ and $||\boldsymbol{\theta}||_1 = 2$.

Solution:



Question 2 (8 marks)

In class, we looked at ridge regression where we minimize the loss function

$$\mathcal{L}_{reg}(\boldsymbol{\theta}) = \frac{1}{2}||(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta})||_2^2 + \frac{\lambda}{2}||\boldsymbol{\theta}||_2^2$$

(a) **[6 marks]** Derive the closed-form solution for the ridge regression problem, that is, find $\boldsymbol{\theta}$ that minimizes the loss function. We require you to show all the steps. The following information can be useful: $\nabla \mathbf{a}^T\mathbf{x} = \mathbf{a}$, $\nabla \mathbf{x}^T\mathbf{x} = 2\mathbf{x}$, and $\nabla \mathbf{x}^T\mathbf{P}\mathbf{x} = 2\mathbf{P}\mathbf{x}$ for a symmetric matrix $\mathbf{P}$.

Solution:
Noting that the loss function is convex, we take its gradient

$$\nabla \mathcal{L}_{reg}(\boldsymbol{\theta}) = \frac{1}{2}\big(-2\boldsymbol{X}^T\boldsymbol{y} + 2\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\theta} + 2\lambda\boldsymbol{\theta}\big)$$

Equating this to zero, we can reach the final solution

$$\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\theta} + \lambda\boldsymbol{\theta} = \boldsymbol{X}^T\boldsymbol{y}$$
$$(\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{I})\boldsymbol{\theta} = \boldsymbol{X}^T\boldsymbol{y}$$

$$\boldsymbol{\theta} = (\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{I})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

(b) **[2 marks]** How does the solution change as $|\lambda| \to 0$ and $|\lambda| \to \infty$? Also identify in which case the model over-fits and under-fits to the data.

Solution:
$|\lambda| \to 0$, we have non-regularized solution, regularized weight vector is the same as the original weight vector.
Overfits to the data.
$|\lambda| \to \infty$, the solution is a zero vector.
Underfits to the data.