

Name: _____

Reg No: _____

Part A: Short Questions (answer in given space)

1. What is the main difference between supervised and unsupervised learning?
2. Why do we use the sigmoid function in logistic regression?
3. What is the key difference between Ridge and Lasso regression?

Part B: Fill in the Blanks

4. In Gradient Descent, the learning rate controls the _____ of the parameter updates.
5. In Lasso Regression, the regularization term is based on the _____ norm of the weight vector.
6. An underfitting model usually has _____ training error and _____ validation error.

Part C: Multiple Choice Questions

7. Which of the following best describes Ridge Regression? (a) Uses L1 norm penalty to shrink coefficients. (b) Uses L2 norm penalty to shrink coefficients. (c) Combines both L1 and L2 penalties. (d) Does not use any regularization.
8. Suppose a model has very low training error but very high validation error. This indicates: (a) Underfitting (b) Overfitting (c) Just right complexity (d) Data is perfectly clean
9. In Elastic Net, setting the mix ratio $r = 1$ makes it equivalent to: (a) Ridge Regression (b) Lasso Regression (c) Linear Regression without regularization (d) Logistic Regression

Part D: Computational Problem

Consider the linear model:

$$h(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

You are given 5 training examples:

i	x_1	x_2	y
1	1	2	10
2	2	1	9
3	0	1	4
4	3	2	15
5	2	3	16

Initial parameters: $\theta_0 = \theta_1 = \theta_2 = 0$. Learning rate: $\alpha = 0.01$.

Hint: The gradient for the j -th parameter is

$$\frac{\partial J}{\partial \theta_j} = \frac{1}{m} \sum_{i=1}^m (h(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

where $x_0^{(i)} = 1$ for all i .

(a) Compute the gradients for all three parameters $(\theta_0, \theta_1, \theta_2)$.

(b) Update the parameters after one step of Gradient Descent.