Name:____ Roll Number:_________________________

Quiz-3

Max. Time: 20 mins Max. Marks: 20

Note: Solve all parts. Limit your written responses to the provided space.

Q.1. [**5×1=5**] Choose by putting a check mark on the most appropriate option or encircling it. Note: No cutting/overwriting is allowed.

1. Fitting a given dataset with a higher degree polynomial is always a good strategy for regression if noise is suspected in the data.

(A) True (B) False

2. Logistic regression for binary classification can only yield linear decision boundaries in the feature space.

(A) True (B) False

3. Stochastic gradient descent is more efficient for very large datasets compared to the batch gradient descent.

(A) True (B) False

4. Mean squared error can be used as a hypothesis for linear regression.

(A) True (B) False

5. Loss function for logistic regression measures the misclassification penalty.

(A) True (B) False

Q.2. [3+4+3+5]

a) Given a training dataset $\mathcal{D}$ with the feature matrix $\mathcal{X} \in \mathbb{R}^{n\times d}$ and a target vector $\mathcal{Y} \in \mathbb{R}^{n\times 1}$, write the expression for optimal parameters to predict the target value using closed form (normal equations) solution.

$\Theta = (\mathcal{X}^T\mathcal{X})^{-1}\mathcal{X}^T\mathcal{Y}$

b) Compute the cost function when the following training vector is presented to the LMS algorithm for linear regression.

$$< \mathbf{x}, \mathbf{t} > = < [2.5\ 1.6\ \ 2.0\ \ -1.4]^T, 5 >$$

Assume the parameter vector is: $\boldsymbol{\theta} = [0.02\ \ -1.2\ \ 0.05\ \ 0.001\ 0.3]^T$

Solution:

$$h_\theta(\mathrm{x}) = \theta^T \mathrm{x} = -1.4714$$

$$\therefore J(\theta) = 0.5 \times (5 - (-1.4714))^2 = 20.94$$

Name:____ Roll Number:_________________________

c) Why is the mean squared error cost function a bad choice for logistic regression?

Ans.

This is because the sigmoid (logistic) function is non-linear in the parameters of the model. The mean squared error cost function based on this hypothesis (i.e., logistic function) becomes non-convex and, therefore, would be a bad choice to find optimal parameter values with gradient descent learning.

d) Write down the cost function and the parameter update rule for logistic regression.

Cost function:

$$\ell(\theta) = \sum_{i=1}^{m} y^{(i)} \log h(x^{(i)}) + (1 - y^{(i)}) \log(1 - h(x^{(i)}))$$

Parameter update rule (using stochastic gradient descent):

$$\theta_j := \theta_j + \alpha \left(y^{(i)} - h_\theta(x^{(i)}) \right) x_j^{(i)} \qquad \text{(for every } j\text{)}$$

Name:____ Roll Number:_________________________