Quiz 4

Graded

Student

HARRIS DOAN

Total Points

14 / 14 pts

Question 1

Linear regression — **4** / 4 pts

1.1 (no title) — **2** / 2 pts

- ✓ **+ 2 pts** Correct

1.2 (no title) — **2** / 2 pts

- ✓ **+ 2 pts** Correct

Question 2

Overfitting and Underfitting — **3** / 3 pts

2.1 (no title) — **1** / 1 pt

- ✓ **+ 1 pt** Correct

2.2 (no title) — **1** / 1 pt

- ✓ **+ 1 pt** Correct

2.3 (no title) — **1** / 1 pt

- ✓ **+ 1 pt** Correct

Question 3

Ridge regression **4** / 4 pts

3.1 (no title) **1** / 1 pt

✓ **+ 1 pt** Correct

3.2 (no title) **1** / 1 pt

✓ **+ 1 pt** Correct

3.3 (no title) **1** / 1 pt

✓ **+ 1 pt** Correct

3.4 (no title) **1** / 1 pt

✓ **+ 1 pt** Correct

Question 4

Regularized Logistic Regression **2** / 2 pts

✓ **+ 2 pts** Correct

Question 5

Transformation of input features **1** / 1 pt

✓ **+ 1 pt** Correct

Q1 Linear regression
4 Points

Consider fitting a linear regression $h_\theta(x) = \theta_0 + \theta_1 x$ to a training dataset $\mathcal{D} = \{(x_1, y_1), ..., (x_n, y_n)\}$, where x_i's are features and y_i's are labels. Let $\mathcal{D}$ consist of 3 data points: $(x_1, y_1) = (1, 0), (x_2, y_2) = (0, 2), (x_3, y_3) = (-1, -1)$.

Q1.1
2 Points

Our current estimate of the parameter is $(\theta_0, \theta_1) = (0, 0)$. Which vector has the same direction as the updated parameter after one step of (batch) gradient descent?

- ○ $[-1, 1]$
- ○ $[-1, -1]$
- ◉ $[1, 1]$
- ○ $[1, -1]$

Q1.2
2 Points

What is the closed form solution for (θ_0, θ_1)?

- ○ $[-1/3, 1/2]$
- ○ $[1/3, -1/2]$
- ○ $[-1/3, -1/2]$
- ◉ $[1/3, 1/2]$

Q2 Overfitting and Underfitting

3 Points

Q2.1

1 Point

If a model is **over-fitted**, then adding new training examples generally decreases the training error.

- (●) False
- ( ) True

Q2.2

1 Point

If a model is **over-fitted**, then adding new training examples generally decreases the test error.

- ( ) False
- (●) True

Q2.3

1 Point

Which approach can generally help prevent underfitting in the corresponding models?

- ( ) Logistic Regression: Add L2 regularization on the weight parameters.
- ( ) Linear regression: Use the closed form solution instead of gradient descent (assuming the step size for gradient descent is set appropriately).
- ( ) Logistic Regression: Use a higher learning rate η in gradient descent.
- (●) Polynomial regression: Use a larger degree for regression.

Q3 Ridge regression
4 Points

Q3.1
1 Point

You are given a dataset with 3000 samples and 300 features. Applying ridge regression to this dataset with $\lambda = 0.2$ **always** has a unique closed-form solution.

- (●) True
- ( ) False

Q3.2
1 Point

You are given a dataset with 300 samples and 3000 features. Applying ridge regression to this dataset with $\lambda = 0.2$ **always** has a unique closed-form solution.

- (●) True
- ( ) False

Q3.3
1 Point

You are given a dataset with 3000 samples and 300 features. Applying ridge regression to this dataset with $\lambda = 0$ (i.e. standard linear regression) **always** has a unique closed-form solution.

- ( ) True
- (●) False

Q3.4
1 Point

You are given a dataset with 300 samples and 3000 features. Applying ridge regression to this dataset with $\lambda = 0$ (i.e. standard linear regression) **always** has a unique closed-form solution.

- (●) False
- ( ) True

Q4 Regularized Logistic Regression
2 Points

Suppose that you are given a binary classification data set $D = \{(x_1, y_1), ..., (x_n, y_n)\}$ where $x_i \in \mathbb{R}, y_i \in \{0, 1\}$. For this data set, we are learning a model $h_{w,b}(x) = \sigma(wx + b)$, where $\sigma(.)$ is the sigmoid function.

Let $J_{logistic}(w, b)$ be the logistic regression cost function and $J_{reg}(w, b) = J_{logistic}(w, b) + \lambda w^2$, $\lambda > 0$ be the regularized logistic regression cost function.

Let $\hat{w}_{logistic}$ be the logistic regression estimator of w and $\hat{w}_{reg}$ be the regularized logistic regression estimator of w. Which of the following statements is **always** true?

- ( ) $|\hat{w}_{reg}| = |\hat{w}_{logistic}|$
- (●) $|\hat{w}_{reg}| \leq |\hat{w}_{logistic}|$
- ( ) $|\hat{w}_{logistic}| \leq |\hat{w}_{reg}|$

($|a|$ denotes absolute value of a).

Q5 Transformation of input features

1 Point

In the following figure, we have a set of data points $X_1, ..., X_N$ where $X_i = (x_i, y_i) \in \mathbb{R}^2$ that are clustered into blue and red classes. We transform X_i's from $\mathbb{R}^2$ to $\mathbb{R}$ by applying the function f defined as :

$$Z_i = f(X_i) = x_i^2 + y_i^2$$

Is the transformed data Z_i's linearly separable?

- (●) Yes
- ( ) No