CSCI 6751 - Past Exam Questions

Quiz #1 and Midterm | With Complete Solutions

Quiz #1 (October 14, 2025)

> **Total Points:** 50 | **Time Allowed:** 40 minutes

Question 1 (25 points) - Gradient Descent

Problem Statement

We train a simple linear regression model: **y = ax + b**

Given data:

x	y
1	3
2	5

Initial parameters: a = 0, b = 0

Learning rate: $\eta = 0.1$

Loss function (MSE):

$$E(a, b) = \frac{1}{n} \sum (y_i - \hat{y}_i)^2$$

where $\hat{y}_i = ax_i + b$ and n = 2

Task: Compute one iteration of Gradient Descent. Show prediction, error, gradients, and updated values of a and b.

Solution

Step 1: Compute Predictions

$$\hat{y}_1 = a \cdot x_1 + b = 0 \cdot 1 + 0 = 0$$

$$\hat{y}_2 = a \cdot x_2 + b = 0 \cdot 2 + 0 = 0$$

Step 2: Calculate Errors

$$e_1 = \hat{y}_1 - y_1 = 0 - 3 = -3$$

$$e_2 = \hat{y}_2 - y_2 = 0 - 5 = -5$$

Step 3: Compute Gradient (∂E/∂a)

$$\frac{\partial E}{\partial a} = \frac{2}{n} \sum (e_i \cdot x_i)$$

$$= \frac{2}{2}[(e_1 \cdot x_1) + (e_2 \cdot x_2)]$$

$$= \frac{2}{2}[(-3 \cdot 1) + (-5 \cdot 2)]$$

$$= \frac{2}{2}[-3 - 10] = \frac{2}{2} \times (-13) = -13$$

Step 4: Compute Gradient (∂E/∂b)

$$\frac{\partial E}{\partial b} = \frac{2}{n} \sum e_i$$

$$= \frac{2}{2}[e_1 + e_2]$$

$$= \frac{2}{2}[(-3) + (-5)]$$

$$= \frac{2}{2} \times (-8) = -8$$

Step 5: Update Parameters

$$a_{new} = a_{old} - \eta \cdot \frac{\partial E}{\partial a}$$

$$= 0 - 0.1 \times (-13) = 0 + 1.3 = 1.3$$

$$b_{new} = b_{old} - \eta \cdot \frac{\partial E}{\partial b}$$

$$= 0 - 0.1 \times (-8) = 0 + 0.8 = 0.8$$

Final Answer:

a_{new} = 1.3, b_{new} = 0.8

New model: $y = 1.3x + 0.8$

Question 2 (25 points) - Fuzzy Logic

Problem Statement

Fuzzy sets for Temperature:

- Low: triangular (0, 0, 25)
- Medium: triangular (20, 30, 40)
- High: triangular (35, 50, 50)

Fan speed outputs:

- Slow = 20
- Medium = 50
- Fast = 80

Rules:

1. IF Temperature is Low THEN Speed is Slow
2. IF Temperature is Medium THEN Speed is Medium
3. IF Temperature is High THEN Speed is Fast

Input: Temperature = 30°C

Tasks:

(a) Compute the degree of membership of Temperature = 30°C in each fuzzy set (Low, Medium, High). Show your calculations using the triangular membership functions.

(b) Using the centroid (weighted average) method, compute the defuzzified fan speed output.

Solution

Part (a): Membership Degrees

Fuzzy Set	Parameters (a, b, c)	μ(T=30)	Calculation

Low	(0, 0, 25)	**0**	$30 > 25$ (c) $\rightarrow 0$
Medium	(20, 30, 40)	**1**	$30 = b$ (peak) $\rightarrow 1$
High	(35, 50, 50)	**0**	$30 < 35$ (a) $\rightarrow 0$

Part (b): Defuzzification Using Centroid Method

Firing Strengths:

Rule	Firing Strength	Output
Rule 1: Low → Slow	0	20
Rule 2: Medium → Medium	1	50
Rule 3: High → Fast	0	80

Centroid Formula:

$$\text{Fan Speed} = \frac{\sum(\text{FS}_i \times \text{Output}_i)}{\sum \text{FS}_i}$$

$$= \frac{(0 \times 20) + (1 \times 50) + (0 \times 80)}{0 + 1 + 0}$$

$$= \frac{0 + 50 + 0}{1} = \frac{50}{1} = 50$$

Final Answer:

Fan Speed = 50

Midterm Examination (October 14, 2025)

> **Total Points:** 50 (Calculation) + 50 (MCQ)
>
> **Time Allowed:** 50 minutes (Calculation) + 15 minutes (MCQ)

Question 1 (25 points) - Multivariate Linear Regression with L2 Regularization

Problem Statement

You are working with a multivariate linear regression model with the hypothesis function:

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

Given:

- A single data point: $(x_1, x_2, y) = (1, 3, 8)$
- Current model parameters: $\theta = [\theta_0, \theta_1, \theta_2] = [1, 0.5, 1]$
- Learning rate for Gradient Descent: $\eta = 0.05$
- L2 Regularization parameter: $\lambda = 0.5$

Tasks:

(i) Using the single data point, perform one iteration of Gradient Descent.

- a) Calculate the prediction $\hat{y}$
- b) Calculate the prediction error
- c) Calculate the updated parameter vector θ_{new} after one GD step

(ii) Using the initial parameters $\theta = [1, 0.5, 1]$, calculate the L2 regularized cost J_{Ridge}. (Assume we do not regularize θ_0).

Solution

Part (i): Gradient Descent

a) Prediction:

$$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$
$$= 1 + 0.5 \times 1 + 1 \times 3$$
$$= 1 + 0.5 + 3 = 4.5$$

b) Error:

$$e = \hat{y} - y = 4.5 - 8 = -3.5$$

c) Gradients (MSE component):

$$\frac{\partial \text{MSE}}{\partial \theta_0} = \frac{2}{n} \sum e_i = 2 \times (-3.5) = -7$$

$$\frac{\partial \text{MSE}}{\partial \theta_1} = \frac{2}{n} \sum (e_i \times x_{1i}) = 2 \times (-3.5) \times 1 = -7$$

$$\frac{\partial \text{MSE}}{\partial \theta_2} = \frac{2}{n} \sum (e_i \times x_{2i}) = 2 \times (-3.5) \times 3 = -21$$

Parameter Updates:

$$\theta_{0,new} = 1 - 0.05 \times (-7) = 1 + 0.35 = 1.35$$

$$\theta_{1,new} = 0.5 - 0.05 \times (-7) = 0.5 + 0.35 = 0.85$$

$$\theta_{2,new} = 1 - 0.05 \times (-21) = 1 + 1.05 = 2.05$$

Part (ii): L2 Regularized Cost

MSE:

$$\text{MSE} = \frac{1}{n} \sum (y_i - \hat{y}_i)^2 = \frac{1}{1}(8 - 4.5)^2 = 3.5^2 = 12.25$$

L2 Penalty (excluding θ_0):

$$\text{L2 Penalty} = \lambda \sum_{j=1}^{2} \theta_j^2 = 0.5 \times (\theta_1^2 + \theta_2^2)$$

$$= 0.5 \times (0.5^2 + 1^2) = 0.5 \times (0.25 + 1) = 0.5 \times 1.25 = 0.625$$

Total Regularized Cost:

$$J_{Ridge} = \text{MSE} + \text{L2 Penalty}$$

$$= 12.25 + 0.625 = 12.875$$

Final Answers:

(i) θ_{new} = [1.35, 0.85, 2.05]

(ii) J_{Ridge} = 12.875

Question 2 (25 points) - Fuzzy Logic System (Coffee Maker)

Problem Statement

System: Smart Coffee Maker Strength Control

Input Variable 1 - Coffee Bean Freshness (days since roast):

- Fresh: trapmf(0, 0, 3, 5)
- Medium: trapmf(3, 5, 10, 14)
- Old: trapmf(10, 14, 21, 21)

Input Variable 2 - Water Quality (ppm minerals):

- Soft: trapmf(0, 0, 50, 100)
- Medium: trapmf(50, 100, 150, 200)
- Hard: trapmf(150, 200, 300, 300)

Output Variable - Brew Strength:

- Mild: trapmf(0, 0, 3, 4)
- Balanced: trapmf(3, 4, 6, 7)
- Strong: trapmf(6, 7, 10, 10)

Rules:

1. IF Beans are Fresh AND Water is Soft THEN Strength is Mild
2. IF Beans are Medium AND Water is Medium THEN Strength is Balanced
3. IF Beans are Old AND Water is Hard THEN Strength is Strong

Current Input:

- Bean Freshness = 6 days
- Water Quality = 120 ppm

Task: Calculate the firing strength of each rule.

Solution

Step 1: Membership Values for Bean Freshness = 6 days

Fresh: trapmf(0, 0, 3, 5)

$6 > d\ (5) \rightarrow \mu_{Fresh}(6) = \mathbf{0}$

Medium: trapmf(3, 5, 10, 14)

6 is in plateau region $[b, c] = [5, 10] \rightarrow \mu_{Medium}(6) = \mathbf{1}$

Old: trapmf(10, 14, 21, 21)

$6 < a\ (10) \rightarrow \mu_{Old}(6) = \mathbf{0}$

Step 2: Membership Values for Water Quality = 120 ppm

Soft: trapmf(0, 0, 50, 100)

$120 > d\ (100) \rightarrow \mu_{Soft}(120) = \mathbf{0}$

Medium: trapmf(50, 100, 150, 200)

120 is in plateau region $[b, c] = [100, 150] \rightarrow \mu_{Medium}(120) = \mathbf{1}$

Hard: trapmf(150, 200, 300, 300)

$120 < a\ (150) \rightarrow \mu_{Hard}(120) = \mathbf{0}$

Step 3: Firing Strengths (Using MIN Operator for AND)

Rule	Calculation	Firing Strength
Rule 1: Fresh ∧ Soft → Mild	min(0, 0)	**0**
Rule 2: Medium ∧ Medium → Balanced	min(1, 1)	**1**
Rule 3: Old ∧ Hard → Strong	min(0, 0)	**0**

Final Answer:

Rule 1 Firing Strength: 0

Rule 2 Firing Strength: 1

Rule 3 Firing Strength: 0

Only Rule 2 is activated.

Midterm - Multiple Choice Questions (Selected)

> **Total Points:** 50 (10 questions × 5 points)
>
> **Time Allowed:** 15 minutes

Question 3 - When to Use Gradient Descent vs. Normal Equation

> The primary reason to use Gradient Descent for linear regression instead of the Least Squares Solution is when:
>
> a. The model is severely underfitting.
>
> b. The number of features is very large (e.g., >1000), making LSS computationally expensive.
>
> c. The relationship between variables is perfectly linear.
>
> d. You need a 100% accurate model.

> **Correct Answer: (b)**
>
> **Explanation:** When the number of features is very large (>1000), computing the matrix inverse $(X^TX)^{-1}$ in the Normal Equation becomes computationally expensive. Gradient Descent avoids this matrix inversion and scales better to high-dimensional problems.

Question 4 - L1 vs. L2 Regularization

> You are building a linear regression model and suspect that only 5 out of 100 features are truly predictive. Which regularization technique would be most appropriate to help identify these key features?
>
> a. L2 Regularization (Ridge)
> b. L1 Regularization (Lasso)
> c. ElasticNet with a higher weight on the L2 part
> d. No regularization is needed.

> **Correct Answer: (b) L1 Regularization (Lasso)**
>
> **Explanation:** L1 regularization (Lasso) drives some coefficients exactly to zero, effectively performing feature selection. L2 regularization (Ridge) only shrinks coefficients but does not set them to zero. For identifying a small subset of important features from many candidates, L1 is the appropriate choice.

Question 6 - Normal Equation Matrix Dimensions

In the multivariate linear regression normal equation $\theta = (X^TX)^{-1}X^TY$, if the design matrix X has dimensions $m \times (n+1)$ (m examples, n features plus intercept), and Y is $m \times 1$, what are the dimensions of the resulting parameter vector θ?

a. $m \times 1$

b. $(n+1) \times 1$

c. $m \times (n+1)$

d. $(n+1) \times m$

Correct Answer: (b) $(n+1) \times 1$

Explanation: The parameter vector θ must have one element for each feature plus the intercept term. Since there are n features plus 1 intercept, θ has dimension $(n+1) \times 1$.

Question 7 - Overfitting and Underfitting Transition

> A model has high error on both training and test data. Increasing model complexity reduces training error to near zero, but test error remains high. This sequence describes the transition from a model suffering primarily from ______ to one suffering primarily from ______.
>
> a. High variance; High bias
> b. High bias; High variance
> c. Underfitting; Optimal fitting
> d. High bias; Low bias

> **Correct Answer: (b) High bias; High variance**
>
> **Explanation:** Initially, high errors on both training and test data indicate underfitting (high bias). After increasing complexity, low training error but high test error indicates overfitting (high variance). The model has transitioned from high bias to high variance.

End of Past Exam Questions