CSCI 6751 V1 | Artificial Intelligence

Midterm Examination

Oct 14, 2025

Total 50 points **Time: 15 minutes**

GOOD LUCK ***Group 2***

Student Name & ID _____

MCQ: Select only the most appropriate option.

Question 1. (5 points)

A model is trained with Gradient Descent. The cost function decreases very slowly and erratically, requiring a huge number of iterations. The learning rate is suspect. What is the most likely scenario?

a) The learning rate is too high; it should be decreased to prevent overshooting.

b) The learning rate is too low; it should be increased to navigate flat regions and saddle points more effectively.

c) The learning rate is optimal; the problem is a non-convex cost function.

d) The model has converged; no action is needed.

Question 2. (5 points)

When using an iterative algorithm like Gradient Descent, stopping the training process before full convergence can act as a form of regularization. This is because:

a) It explicitly adds a penalty term to the loss function, similar to L2 regularization.

b) It limits the effective complexity of the model by preventing it from perfectly fitting the training noise, thus reducing variance.

c) It ensures the model remains linear, thus increasing bias but improving interpretability.

d) It forces the parameters to be exactly zero, mimicking L1 regularization.

Question 3. (5 points)

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

a) m * 1 b) (n+1) * 1 c) m * (n+1) d) (n+1) * m

Question 4. (5 points)

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

Question 5. (5 points)

Given a rule "If X is A and Y is B then Z is C," with input membership values μ_A(x)=0.7 and μ_B(y)=0.8, the firing strength is:

a) 1.5 b) 0.56 c) 0.7 d) 0.8

Question 6. (5 points)

The expected prediction error for a new test point x_0 can be decomposed as $E[(y_0 - \hat{y}_0)^2] = Var(\hat{y}_0) + [Bias(\hat{y}_0)]^2 + Var(\varepsilon)$.

If you take a model with high variance and add a very strong L2 regularization penalty (large λ in Ridge), what is the primary effect on this decomposition?

a) Variance increases and Bias decreases.

b) Variance decreases and Bias increases.

c) Both Variance and Bias decrease.

d) The irreducible error, Var(ε), is reduced.

Question 7. (5 points)

A = {(x1, 0.4), (x2, 0.7)}

B = {(x1, 0.8), (x2, 0.2)}

What is the membership value for element x1 in the union A∪B, defined as μ_A∪B(x) = max[μ_A(x), μ_B(x)]?

a) 0.4 b) 0.8 c) 0.32 d) 1.2

Question 8. (5 points)

Given a rule "If X is A and Y is B then Z is C," with input membership values μ_A(x)=0.7 and μ_B(y)=0.8, the firing strength is:

a) 1.5 b) 0.56 c) 0.7 d) 0.8

Question 9. (5 points)

What is the key difference between the "support" and the "core" of a fuzzy set?

a) The support has $\mu(x) > 0$, the core has $\mu(x) = 1$.

b) The support has $\mu(x) = 1$, the core has $\mu(x) > 0$.

c) The support is always a subset of the core.

d) They are two terms for the same concept.

Question 10. (5 points)

During Gradient Descent, if the learning rate is set too high, what is a likely consequence?

a) The algorithm will converge too slowly.

b) The algorithm may overshoot the minimum and fail to converge, or even diverge.

c) The algorithm is guaranteed to find the global minimum faster.

d) The model will inevitably underfit.