DS 4400 Spring 2026: Practice Exam

February 23, 2026

Time: 100 minutes

NAME (PRINTED): _____

Show all your work and calculations. Provide justification for your answers. Partial credit will be given for work that is partially correct. Points will be deducted for false statements, even if the final answer is correct.

You can bring one printed page with notes from the class lectures and a calculator to the exam. You are not allowed to consult any online resources or discuss with your colleagues.

Problems

1. Name a supervised learning model that can handle multi-class classification natively (without modification).
2. Explain the difference between training error and test error. Why is test error typically higher? What does it mean if training error is much lower than test error?
3. What is cross-validation and why is it useful? Name one limitation of using cross-validation.
4. In k-NN classification, what happens to the decision boundary as k increases? What about as k approaches the number of training samples?
5. Explain why a more complex model is not always better, even if it has lower training error.
6. You have data that looks like the following

   $$\begin{bmatrix} y & x_1 & x_2 & x_3 & x_4 & x_6 & x_7 & x_8 \\ \text{Rent} & \text{\# Bedrooms} & \text{Sq. Ft.} & \text{Sq. m.} & \text{\# Windows} & \text{Zipcode} & \text{Electric Heating (Y/N)} & \text{Gas Heating (Y/N} \end{bmatrix}$$

   You have data for 10000 apartments. What algorithm would you use? How would you optimize this algorithm? Justify why/why not for the optimization method you pick. What type of feature pre-processing or normalization would you use?
7. Case study: Customer Churn Prediction

   A telecommunications company wants to predict which customers are likely to cancel their service (churn) in the next month. Available data includes customer demographics, service usage patterns, billing history, and customer service interactions. Describe:

   - Relevant features and any preprocessing needed. What additional data would you add?
   - Two appropriate machine learning algorithms
   - Two evaluation metrics
8. As we studied in class, Linear Regression and can be trained either by Gradient Descent or directly by computing the closed-form solution. List at least two limitations of Gradient Descent compared to the closed-form solution.
9. You are running gradient descent and you notice that your learned parameters lie in very different scales. How would you modify your gradient descent equation to account for this? Write down the relevant equations and explain how they help alleviate the problem.
10. Given some θ_0, θ_1 and features for a new data point x_1, x_2 and an assumed threshold of 0.5, can you predict if the new point is going to be classified as positive or negative without access to the sigmoid function?
11. Write out the confusion matrix - then write out the formulas for precision and recall in terms of TP, FP, FN and TN.
12. Give an example of a situation where you care more about precision than recall.