CS461 Quiz1

September 24, 2025

Name: _____

Instructions: Mark your answers directly on the test, with an x or checkmark on your chosen solutions. Multiple choice questions are 15% each, and the last question is 25%.

1) Which of the following statements are **true**?

a. Unbiased estimators have the lowest error of any estimators.
b. The sin function is not polynomial so bias cannot be measured.
c. 0 is an unbiased estimator of sin(x)
d. The sin function exists from $-\infty$ to ∞ so bias cannot be measured.

2) As the amount of noise increases, which of the following is **false**?

a. More data is required to estimate the signal.
b. More computationally expensive machine learning techniques are required.
c. The quality of the signal estimate worsens.
d. If the noise is i.i.d., then the mean of the samples is still the best estimate of the signal's mean.

3) Which of these sentences best describes the bias variance trade-off?

a. The person modelling should choose a bias-variance trade-off based upon knowledge of the dataset.
b. A model should capture the variance in a dataset by increasing its complexity and reducing bias to a presumed distribution.
c. A model should be simplified to bias it to avoid matching the variance of a dataset.
d. The bias-variance trade-off is inherent to a dataset and is not controlled by the model.

4) When we modify the normal equation like this, $\beta = (X^T X + \lambda I)^{-1} X^T y$, the λ

a. is a meaningless constant.
b. is only there to stabilize the matrix inversion.
c. shapes the parameters, biasing the model to better match noise of a guassian distribution.
d. is a heuristic that was invented by Laplace to make Guass' least squares regression work on non-Guassian data.

5) Which of the following is **false** about decision boundaries?

a. For samples on the decision boundary, we expect the probability of belonging to a class to be near 0.5.
b. Logistic regression will create decision boundaries based upon sample statistics, not prediction error rates.
c. The perceptron model will create decision boundaries based upon sample statistics, not prediction error rates.
d. The sigmoid function converts a decision boundary to a probability estimate.

6) Build a decision tree from the following data, predicting the class from x and y:

class	x	y
a	0	0
a	1	1
a	8	8
b	2	1
b	1	2
b	6	6

Use the Gini impurity, $1 - \sum_{c=1}^{C} \hat{p}_c^2$. At each node on the tree, write either the pivot value or, for the leaf nodes, the class. If multiple pivots have the same Gini impurity, you may choose any one of them arbitrarily. Splits are $<$ and $\geq$. You do not need to calculate every value; try plotting them if you are confused.