CS 5703 Machine Learning Practice
Midterm 1
October 8, 2026

General instructions:

- Please wait to open this exam booklet until you are told to do so.
- This examination booklet has 8 pages.
- Write your name and the date, and sign your name below.
- The exam is closed book, notes, and electronic devices. The exception is that you may have one page of personal notes (double sided).
- The exam is worth a total of 150 points (and 15% of your final grade).
- You have 1.25 hours to complete the exam. Be a smart test taker: if you get stuck on one problem, go on to the next.

On my honor, I affirm that I have neither given nor received inappropriate aid in the completion of this exam.

Signature: _____ **Name:** _____

Date: _____

Question	Points	Score
Classifiers	60	
Regression	22	
Convolution	25	
Pipelines	24	
Metrics and Gradients	19	
Total:	150	

Part I. Classifiers

1. (8 points) Explain the reason why *accuracy* is non-ideal for use as a loss function.

2. (10 points) Assume that a learned decision function $f(.)$ produces the values for the following samples, and assume a threshold of 0.4.

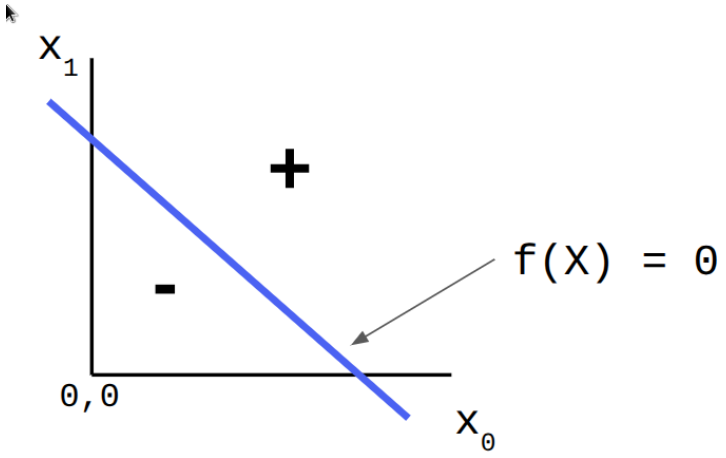
sample	**True Label**	$f(.)$
0	F	$-.23$
1	F	.2
2	T	.72
3	T	$-.3$
4	F	.53
5	F	.37
6	F	$-.5$
7	T	$-.14$

Compute the following:

- TP =
- FP =
- TN =
- FN =

3. (5 points) What is the True Positive Rate? Express as a fraction.

4. (6 points) Consider the following decision boundary.

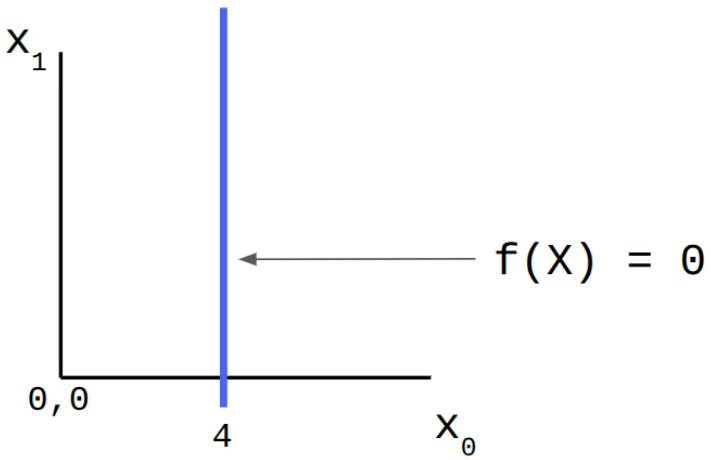


where: $f(X) = w_0x_0 + w_1x_1 + w_2$

Which is true?
A. $w_2 < 0$ B. $w_2 = 0$ C. $w_2 > 0$

5. (6 points) For the above decision boundary, which is true?
A. $w_0 < 0$ B. $w_0 = 0$ C. $w_0 > 0$

6. (6 points) Consider the following decision boundary; assume that $w_2 = 8$.



What is the value of w_0?
$w_0 =$

7. (6 points) Which side of the decision line is the positive (true) region?
A. $x_0 < 4$ B. $x_0 > 4$

8. (5 points) What is the range of possible Area Under the Curve (AUC) values?

9. (8 points) Under what conditions will a classifier achieve a high AUC? Explain in terms of the distributions of the classifier scores.

Part II. Regression

Consider the following set of samples, the true labels, and the regression output from two different models:

Sample Index	**True Value**	$f_1(X)$	$f_2(X)$
0	3	4	2
1	7	5	4
2	4	2	4
3	2	5	5

10. (8 points) What is the Mean Squared Error for model 1? Express as a simplified fraction.

    $MSE_1 =$

11. (8 points) What is the Mean Absolute Error for model 2? Express as a simplified fraction.

    $MAE_2 =$

12. (6 points) Which model is best? Justify your answer.

Part III. Convolution

Consider the following vector:

X = [0 0 0 1 0 0 1 1 0 0 0 1 1 1 0 0 0]

and a kernel: K = [-1 1 1 -1]

13. (10 points) What is $K * X$? Assume no padding.

14. (5 points) What is the highest value in the sequence? Explain why.

15. (10 points) Assume a new kernel $\hat{K}$ of length 3 and the same X as above.
$\hat{K} * X$ = [0 -1 0 1 -1 -1 1 1 0 -1 -1 0 1 1 0]

What is $\hat{K}$?

Part IV. Pipelines

16. (6 points) A feature value is a string that is expected to be one of several values: "yes", "no", or "maybe"; however, the string can actually take on any value. Show a reasonable *one-hot* encoding of the possible values.

17. (6 points) List two differences between a DataFrame and a Numpy Array.

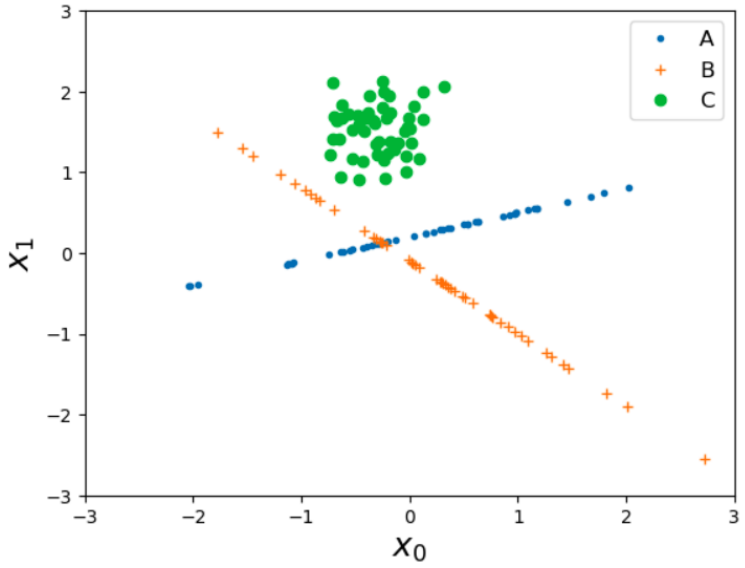
18. (6 points) For the Gaussian convolutional filter, explain the effect of changing the standard deviation.

19. (6 points) Explain why it is problematic to follow our DataDropper pipeline element (removes rows that contains NaNs) with our implementation of the Derivative pipeline element.

Part V. Metrics and Gradients

20. (5 points) What is the range of the Pearson correlation metric?

21. (6 points) Given the following sets of samples in a 2D grid:



Estimate Pearson's Correlation between x_0 and x_1 for each of sets A, B, and C.

A :

B :

C :

22. (8 points) In Logistic Regression with a MSE loss function, we showed that:

$$\frac{\delta E_i}{\delta w_j} = -2\left(y_i - g(f(X_i))\right) \times g'(f(X_i)) \times x_{i,j} \ ,$$

where i is the sample index and j is the parameter index. Explain the three factors that may lead to no change in w_j when performing gradient descent with respect to sample i.