CPTS 437 - Introduction to Machine Learning

Spring 2020

Exam #2

Full Name:

WSU ID:

**Duration:** 50 minutes

**Instructions**: Clearly write your name at the top of each page of this exam. Complete all problems on this exam. Write all you work on this exam; you may use the backs of pages if needed. You may use your own calculator. You may *not* use a computer. Failure to turn in your exam at the end of 50 minutes will result in deduction of points. Anyone cheating on the exam will receive a zero.

|  |  |  |
| --- | --- | --- |
| Problem | Points Possible | Your Score |
| 1 | 7 |  |
| 2 | 7 |  |
| 3 | 15 |  |
| 4 | 12 |  |
| 5 | 7 |  |
| 6 | 12 |  |
| 7 | 10 |  |
| 8 | 18 |  |
| 9 | 12 |  |
| Total | 100 |  |

1. (7 points) Which of the following evaluation metrics cannot be applied to logistic regression and why? ***AUC-ROC, Accuracy, Log loss, Mean squared error***

AUC-ROC cannot be applied because the logistic regression function deals with values of 0 and 1 that would not be plotted well with these curves.

2. (7 points) Explain the inductive biases for the perceptron classifier.

There are two main ones. The first being that the model prefers smaller weights so the perceptron doesn’t bounce around as much and can converge easily to the correct weights and bias. The other inductive bias for perceptron classifiers is that if two known similar points lie on a smooth line and a third point is introduced between them on that line, then it will be classified along with the previous two points because it is highly likely that it belongs to the same class.

3. (15 points) Describe the following components of a ranking classifier.

a) (5 points) The input to the binary classifier.

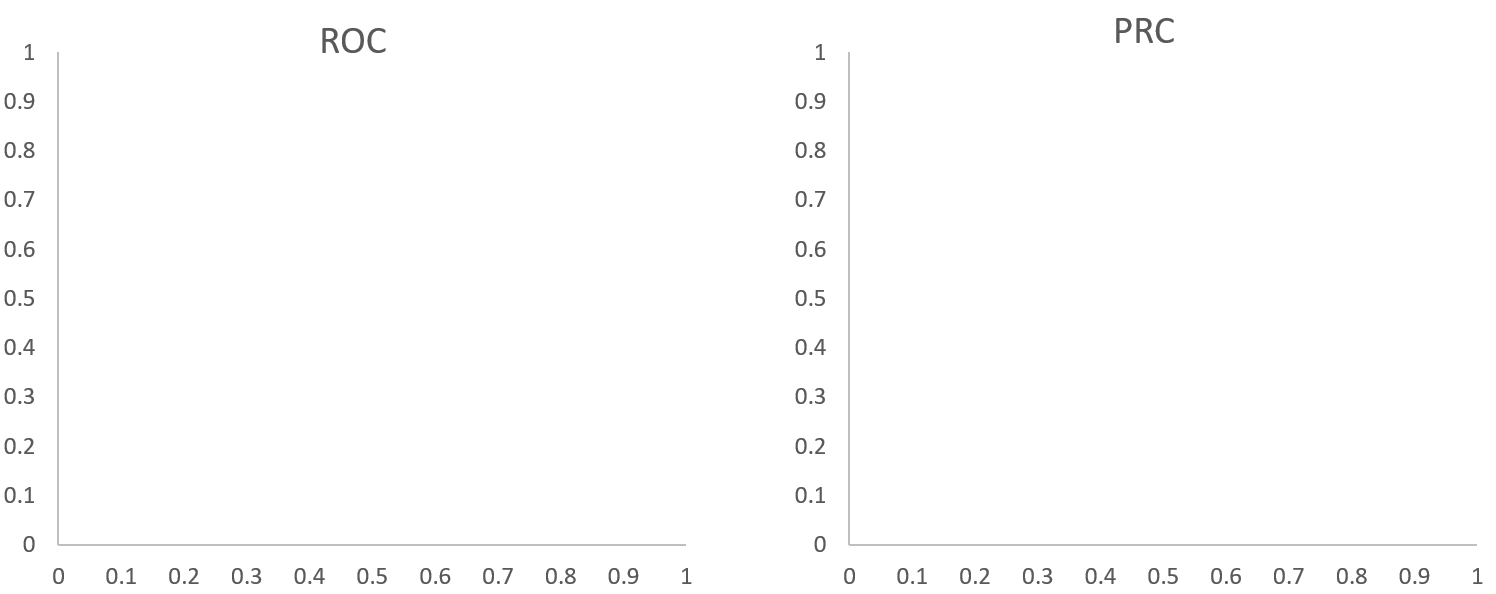
could be a search string or query of some sort for the classifier to find and rank documents based on. Similar to how google uses a search query

b) (5 points) The output of the testing algorithm.

The top-ranked documents, again similar to google’s search results and how the top results are the most matched and highest ranked based on the search query.

c) (5 points) An example loss function.

could be cost function, where the cost of correctly or incorrectly classifying is determined and calculated

4. (12 points) A series of 3 classifiers (A, B, and C) are developed to diagnose a disease. Draw example ROC and PRC curves that reflect the fact that A performs best on this problem, followed by B and C (which performs the worst). Label your plots with the names of the classifiers as well as the names of the axes.

C

B

A

C

B

A

Precision

True Positive Rates

False Positive Rate

Recall

5. (7 points) Consider the set of training points plotted in the graph below. On the graph, draw the decision boundary that would be formed by a perceptron classifier.

A close up of a logo

Description automatically generated

6. (12 points) Consider a situation in which we generated 8 sets of performance values from Algorithms A and B. We want to determine if the difference in performance is statistically significant. A paired t-test returns a t value of 2.9. Would the difference be considered statistically significant? Extremely statistically significant? Provide a brief justification for each of your answers.

A close up of a window

Description automatically generated

Because we generate 8 sets of performance values, our degrees of freedom is 8-1 = 7. Comparing our t-value, we see that 2.9 is between 2.365 and 2.998, thus our p-value will be somewhere between .05 and .02. Thus, our difference would in fact be considered statistically significant because it is less than 0.05 however it is not quite extremely statistically significant. We still have enough evidence to reject the null hypothesis.

7. (10 points) Provide an explanation of why the L1 norm (1-norm) is less susceptible to outliers than the L2 norm (2-norm).

The L1 norm is less susceptible to outlier because it normalizes all weights of the values to either 0 or 1 which will in turn reduce the effect of an outlier for a specific feature that may have a very large value compared to the others (salary was a good outlier example in class with CEO’s). The L2 norm spreads everything out equally for the weights so that one outlying feature value could throw off the weights for the norm.

8. (18 points) Suppose you are analyzing regression data generalized by a polynomial function of degree 3. Characterize the bias and variance of the estimates provided by the following models on the data by writing “low” or “high” in each empty box. Provide short justifications for your answers.

|  |  |  |
| --- | --- | --- |
| ***Estimator*** | ***Bias*** | ***Variance*** |
| Linear regression | High – will not accurately predict many points | Low – will be a consistent line |
| Polynomial regression with degree 3 | Low – fits the model well, accurate predictions of points | Low – fits the model well still, will be consistent |
| Polynomial regression with degree 10 | Low – fits the model somewhat, but will have close to accurate predictions so bias is still low | High – line will try to fit too many points and won’t be consistent |

9. (12 points) Given all the previous patients a doctor has seen (below are their symptoms and diagnosis)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| chills | runny nose | headache | fever | flu |
| yes | no | mild | yes | no |
| yes | yes | no | no | yes |
| yes | no | strong | yes | yes |
| no | yes | mild | yes | yes |
| no | no | no | no | no |
| no | yes | strong | yes | yes |
| no | yes | strong | no | no |
| yes | yes | mild | yes | yes |

Use the naïve Bayes classifier to determine whether the following patient has the flu. Show all your work.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| chills | runny nose | headache | fever | flu |
| yes | no | mild | yes | ? |

