CPTS 437 - Introduction to Machine Learning

Spring 2020

Exam #3

Full Name:

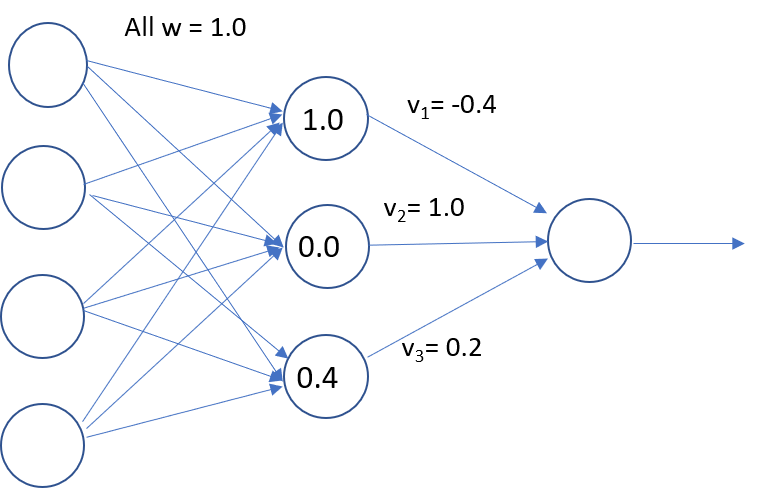
WSU ID:

**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 if needed. 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 | 14 |  |
| 2 | 10 |  |
| 3 | 12 |  |
| 4 | 8 |  |
| 5 | 8 |  |
| 6 | 18 |  |
| 7 | 25 |  |
| 8 | 5 |  |
| Total | 100 |  |

1. (14 points) The figure below shows the structure of a two-layer neural network with the corresponding edge weights. Compute the output of the network for the input (0, 1, 1, 0). Show each step in your calculation. You may leave the final value in the form of a formula if you do not want to perform the calculation. You may also use the tanh values that most closely fits your needs: tanh(0.0)=0.0, tanh(0.5)=0.46, tanh(1.0)=0.76, tanh(1.5)=0.91, tanh(2.0)=0.96.



Input = (0, 1, 1, 0)

(0x1)+(1x1)+(1x1)+(0x1) = 2

h(top) = tanh(2) = 0.96

(0x1)+(1x1)+(1x1)+(0x1) = 2

h(middle) = tanh(2) = 0.96

(0x1)+(1x1)+(1x1)+(0x1) = 2

h(bottom) = tanh(2) = 0.96

Output = (.964 x -.4) + (.964 x 1.0) + (.964 x .2) = .768

1. (10 points) PCA reduces the dimensionality of data by minimizing projection error or equivalently maximizing variance. Plot the directions of the first PCA component in the left figure below. Then plot the direction of the second PCA component in the right figure below.

A screen shot of a computer

Description automatically generatedA screen shot of a computer

Description automatically generated PCA Component 1 PCA Component 2

1. (12 points) What are two ways to establish needed diversity for ensemble classifiers?

Since ensemble methods create multiple models to get better results, taking advantage of things such as voting classifiers and choosing data points while resampling will yield greater diversity and allow the classifier to choose from many models.

1. (8 points) For a neural network, which one of these parameters most affects the trade-off between underfitting and overfitting? Circle the best answer and provide a brief justification.

* number of hidden nodes
* learning rate
* initial choice of weights
* inclusion of bias term

The learning rate will affect the model the most in terms of overfitting or underfitting. Giving a model a learning rate that is too high will cause it to “jump” back and forth over the global min and will likely overfit. On the flip side, choosing a very small learning rate will cause the model to converge very slowly, likely underfitting the model.

All of the above parameters will have an affect on the model in terms of how well it will fit, but choosing the learning rate not only affects this outcome the most, but choosing a poor learning rate can actually affect the performance of the other parameters as well. Your choice of initial weights or bias term will mean very little if your learning rate is very poor to begin with.

1. (8 points) What is meant by a hard margin for an SVM?

A hard margin SVM will only work with linearly separable data and will seek to find a hyperplane that classifies all data correctly. A hard margin is much more susceptible to noisy data

What is the role of the C parameter in the SVM objective function?

The C parameter defines “how soft” the margin should be. The larger C is, the more strict the margin will be for getting every data point correctly classified.

1. (18 points) Compare and contrast linear regression with logistic regression.
   1. (6 points) What is the loss (cost) function for linear regression?

The loss function for linear regression is for minimizing the MSE

* 1. (6 points) What is the loss (cost) function for logistic regression?

The loss function for logistic regression is used for minimizing the log function

* 1. (6 points) Why does logistic regression not use the same loss function as linear regression?

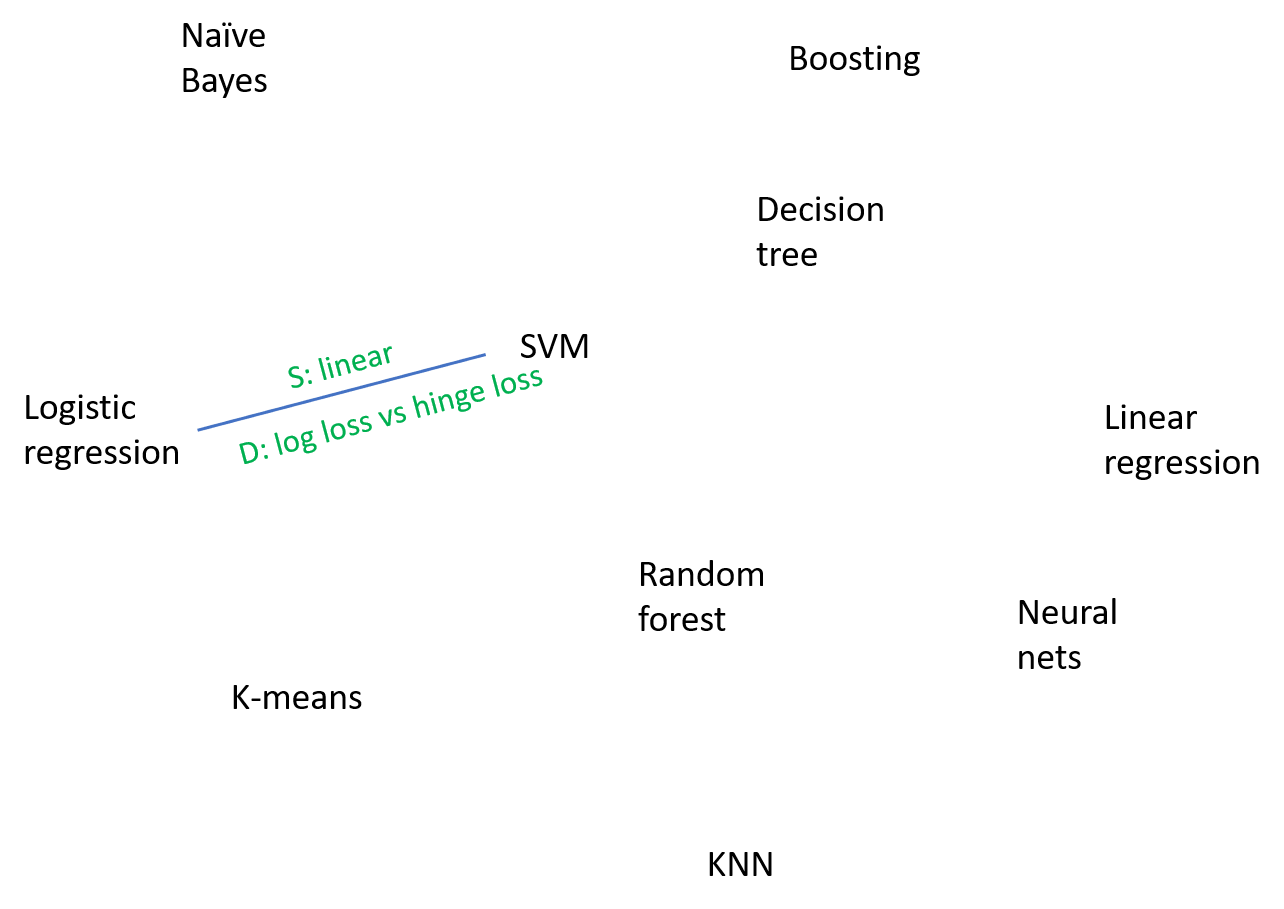
You cannot use the same loss function for logistic and linear regression because linear regression models data using continuous values and will develop a relationship between the feature values and the labels. The logistic regression is used for more of a classification of variables into binary types, usually 1/0 or T/F etc…

Trying to use MSE on logistic function would result in the function never being able to converge as it would not be convex anywhere and thus could not find a minimum.

1. (25 points) Following the example that is given, add 5 edges to the figure below relating a pair of algorithms. Each edge should be labeled with one characteristic the methods share (label as “S:”) and one difference (labeled as “D:”). These labels should be short and address basic concepts such as type of learning problem, loss function, and hypothesis space. An example label is provided for you.

S: boost alg performance

D: RF can use bagging and boosting, but boosting is a single ensemble method



S: clustering algs

D: closest neighbor vs mean clusters

S: Both classification

D: learning func vs tree

S: linear

D: log loss vs MSE

S: Both Trees

D: RF can create trees with separate featrure values and compare

1. (5 points) What is one machine learning project you would like to pursue if you have unlimited time and resources available?

The causes of deadly illnesses that have little to no cure such as cancer, Alzheimer’s, etc…

I’ve heard so many different theories over my entire lifetime about random things that may cause life threatening illnesses later in life, especially mentioning cancer. I’ve heard things from drinking milk past a certain age, eating red meat, or using electronics and cell phones too much can actually cause cancer later in life. This is such a deadly and terrible thing that people seem to be facing more and more and I would really like to get a handle on what causes this or maybe how we can avoid it. But I would want the unlimited resources for this because I am tired of just hearing theories, I want to know for sure that we can stop terrible things like cancer from happening in the first place.