**CS 472 – Introduction to Machine Learning**

**Winter 2022 Midterm Exam**

**Take Home**

**[C. Giraud-Carrier, 2232 TMCB]**

**Open Notes/Book**

1. (10 points) Circle the correct answer.

* T or F : If a decision tree is an elaboration of another decision tree , then is more general than (Def: is an elaboration of if ID3 can extend into ).
* T or F : ID3 cannot learn the XOR function.
* T or F : The hidden layer of a neural network guarantees that Backpropagation does not overfit the training data.
* T or F : Learning is impossible in the absence of bias.
* T or F : The first 3 principal components in PCA always account for over 75% of the variance in the data.
* T or F : Doing machine learning without business understanding is like calling a plumber to fix your car.
* T or F : Perceptrons can solve only linearly separable classification problems.
* T or F : NFL states that on average across all tasks all learning algorithms perform the same.
* T or F : Business users have little to do with the success of a machine learning project.
* T or F : Logistic regression is linear regression in the logit space.

2. (1 point) Complete the sentence: Given that algorithm performs better than Algorithm on a selection of 10 learning tasks, I can rightfully conclude that will outperform …

1. On most learning tasks
2. On some learning tasks
3. On all learning tasks
4. On no learning tasks

3. (2 points) Consider the following set of points:

What simple linear function of in terms of may serve as a reasonable model for ?

What is the RMSE of that model?

1. 0.0145
2. 0.0216
3. 0.0043
4. 0.0657

4. (1 point) A learning algorithm overfits the training data if…

1. has high accuracy on training data and poor accuracy on test data
2. has high accuracy on training data and high accuracy on test data
3. has poor accuracy on training data and high accuracy on test data
4. is too large to fit in memory
5. has poor accuracy on training data and poor accuracy on test data

5. (1 point) Assume that the units of a feedforward neural network are modified so that they compute the function instead of the sigmoid function. Given that the derivative of the function is , what is the resulting backpropagation weight update rule () for the output layer?

6. (1 point) Assume that PCA produces the following eigenvalues:

1.4, 0.45, 2.0, 0.55 and 0.6.

How much of the overall variance is explained by the first 3 principal components?

1. 50%
2. 77%
3. 60%
4. 80%

7. (4 points) Consider the following simple classification learning algorithm, called SingleClass, and answer the associated questions.

current\_prediction ← *unknown*

For each new training example *E*

If current\_prediction != *E*’s target value

current\_prediction ← *E*’s target value

1. Given a sequence of training examples , what is the model produced by this learning algorithm?
2. Assume that the training set is , where the last entry is the target (i.e., , }. What would the algorithm predict for the test example ?

Now, consider the majority learner, i.e., the learning algorithm whose prediction is the most frequently occurring target value among the training examples. Answer the following questions.

1. What would this algorithm predict for the test example ?
2. T or F : The majority learner will always do better than the SingleClass algorithm.

8. (2 points) “For decades, social scientists have been comparing the predictive accuracies of Super Crunchers [or machine learning techniques] and traditional experts. In study after study, there is a strong tendency for the Super Crunchers to come out on top….It’s best to have the man and machine in dialogue with each other, but, when the two disagree, it’s usually better to give the ultimate decision to the statistical prediction….This is in many ways a depressing story for the role of flesh-and-blood people in making decisions….What, if anything, in the process of prediction can we humans do better than the machines?” (Ian Ayres). Provide your own answer to Ian Ayres’ question.

9. (2 points) The Universal Approximation Theorem (UAT) states that any arbitrary function may be approximated to any arbitrary degree of accuracy by some neural network with at most 2 layers of weights. Explain why this does not contradict the No Free Lunch Theorem (i.e., that there is no universal learner) with respect to Backpropagation.

10. (1 point) What is learning bias?

1. The thing you put around the hem of your skirt
2. The thing you use to select one generalization over another
3. The thing you create when you carelessly sample data
4. The thing you apply when you memorize information

11. (2 points) Tom Khabaza (a leading UK data mining consultant) once said: “Projects never fail due to lack of patterns.” If this is true, then what may cause machine learning projects to fail.

12. (2 points) Consider the problem of overfitting.

a) What is a practical way of detecting overfitting?

b) What are two ways to avoid it in decision tree learning?

13. (1 point) A study by a major metropolitan newspaper found that in certain work environments, people who smoke cigarettes are less likely to develop carpal tunnel syndrome (CTS).

1. This is explained by the fact that smokers take more breaks
2. This is explained by the fact that Nicotine inhibits the CTS-inducing gene
3. This is explained by the fact that non-smokers also have poorer postures
4. All of these explanations are speculative
5. None of these explanations is correct

14. (2 points) Find an article in the news over the past couple of years that talks about machine learning (success, failure, speculation, apocalyptic, etc.). Include the link below, and write a couple of paragraphs about what you learned, whether you agree or not (and why), issues you saw, insights you gained, etc.

Link:

Comment: