Name: Gregory Knapp

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 D_2 is an elaboration of another decision tree D_1, then D_1 is more general than D_2 (Def: D_2 is an elaboration of D_1 if ID3 can extend D_1 into D_2).
- 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.

  This question is a little confusing. The intro slide says "if a learning system is to be useful, it must have some form of bias" or else it just memorizes. Useful is not the same as impossible, but I think thats the message this question is trying to get accross so I said True
- 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 A performs better than Algorithm B on a selection of 10 learning tasks, I can rightfully conclude that A will outperform B…

i. On most learning tasks
ii. On some learning tasks
iii. On all learning tasks
iv. On no learning tasks

Name: ______________________________

3. (2 points) Consider the following set of $< x, y >$ points:

$$P = \{< 1.45, 3.97 >, < 6.02, 12.98 >, < 4.55, 10.01 >, < 7.21, 15.43 >, < 2.01, 4.95 >\}$$

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

i. $y = 2.5x$
ii. $y = x + 2$
iii. $y = 2x + 1$
iv. $y = x - 1$

What is the RMSE of that model?

i. 0.0145
ii. 0.0216
iii. 0.0043
iv. 0.0657

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

i. A has high accuracy on training data and poor accuracy on test data
ii. A has high accuracy on training data and high accuracy on test data
iii. A has poor accuracy on training data and high accuracy on test data
iv. A is too large to fit in memory
v. A 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 $tanh$ function instead of the sigmoid function. Given that the derivative of the $tanh$ function is $tanh'(x) = 1 - tanh^2(x)$, what is the resulting backpropagation weight update rule (Δw) for the output layer?

i. $\Delta w = c(t_q - o_q)o_q(1 - o_q)o_p$
ii. $\Delta w = c(t_q - o_q)^2 o_q(1 - o_q)o_p$
iii. $\Delta w = c(t_q - o_q)(1 - o_q^2)o_p$
iv. $\Delta w = c(t_q - o_q)o_q(1 - o_q)(1 - o_p)^2$

Sigmoid — Derivative term

$c(t_q - o_q)$ $\boxed{(1 - o_q)o_p}$

$d_{out} = (t - o_q)\, o_q(1 - o_q)$

$\Delta w = lr \cdot d_{out} * z_{in}$

$d_{out} = (t_q - o_q)(1 - o_q^2)$

$c(t_q - o_q)(1 - o_q^2)o_p$

lr — error — tanh deriv — output of prior node

Name: ______________________________

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

1.4, 0.45, 2.0, 0.55 and 0.6 $\lambda_1 = 2.0$ $\lambda_2 = 1.4$, $\lambda_3 = .6$

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

$$\frac{2.0 + 1.4 + .6}{2.0 + 1.4 + .6 + .45 + .55} = \frac{4}{5} = 80\%$$

i. 50%
ii. 77%
iii. 60%
iv. 80%

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

current_prediction ← *unknown* None
For each new training example *E* E_i
  If current_prediction != *E*'s target value False
    current_prediction ← *E*'s target value $= E_{i,target}$

a) Given a sequence of training examples $S = \{E_1, E_2, \ldots, E_n\}$, what is the model produced by this learning algorithm?

$Model(E_n) = E_{n-1,target}$

A model that always predicts the class of the previous example it found, unless it was a match. It essentially expects a Single class, the one it previously encountered

b) Assume that the training set is $S = \{< red, square, pretty >, < red, triangle, pretty >, < blue, square, ugly >, < red, circle, pretty >, < blue, triangle, ugly >\}$, where the last entry is the target (i.e., $pretty, ugly\}$. What would the algorithm predict for the test example $< red, square, ? >$? After Δ loop → Current_pred = ugly

i. *pretty*
ii. *ugly*

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.

c) What would this algorithm predict for the test example $< red, square, ? >$?

i. *pretty*
ii. *ugly*

pretty 3 ~~ugly 2~~

d) T or F : The majority learner will always do better than the SingleClass algorithm.

True

Majority learner will at worst get 3/5 with 2 classes, Single class loses 2 predictions/change in target class. It can only at best tie the majority learner

Name: ______________________________

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.

Reading the quote above & from my experience, the area where we cannot compete with machines is pure computational power. When a learning algorithm outputs a decision, it has considered many times more data at speeds exponentially faster than humans could. In that regard, there is no competition. However, the area that we still have a key advantage in is in our ability to reason about outcomes, especially when they don't match our expectations. Outliers exist in almost all datasets & while more complex deep learning can do well in adjusting to them, there are cases when the reasoning that a human can provide can help clarify the outcome. The machine only knows the inputs & outputs, but does not need to know the data gathering process to decide. As humans I think we still retain that ability to think critically which is not always there when a machine is just trying to best approximate a function. Our input is still valuable, especially when facing exceptional cases, or high pressure decisions with slim error margins (airline safety, military missions, etc.). ML tends to be most successful when done in coordination with experts in the field of prediction.

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.

The No Free Lunch theorem states there is no universal learner; or in other words, models that perform well on one task will struggle on others. This does not contradict with the UAT, because the UAT insists that there exists at least 1 model that can approximate a function, not one to approximate all functions (1 to 1 not 1 to all). In regards to Backprop, the same is true. If one Backprop Multilayer perceptron works well to approximate certain functions & data sets but struggles with another, that is okay. A separate model (with different hyperparameters) may be able to succeed in the areas where the first model struggled. Both theorems hold. In this case; one model is not universally successful, but the existence of infinite models ensure that one can approximate the function with the right construction & set up.

Name: ______________________________

10. (1 point) What is learning bias?

i. The thing you put around the hem of your skirt
ii. The thing you use to select one generalization over another
iii. The thing you create when you carelessly sample data
iv. 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.

There are a few reasons why ML projects often fail:

- Misapplication of ML: Try as it will, a machine learning algorithm cannot create results out of thin air. If data with no relation to outputs is thrown at a model, the results will not help in any way. Data curation & management is a huge stumbling block, especially when starting a new project.

- Lack of critical analysis of outputs - Like the first cause a working model is great, but does nothing for the project if its outputs are not used intelligently. A business using a ML model will almost surely fail if it simply lets the machine decide everything. Human input, especially from experts of a field, to help analyze model outputs gives greater chances for success in using ML.

- Misplaced Expectations - Sometimes, ML projects are deployed with the expectation that it will discover some new insight in data to put them over the top, when in reality, often ML models help to reinforce & refine already existing & successful processes. If a model is deployed with incorrect expectations, a project will surely fail because it will most likely not accomplish the exact task you set out to do.

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

a) What is a practical way of detecting overfitting?

Looking at training / test or validation loss or accuracy over epochs is a great way to detect overfitting. If a point is reached where validation loss/accuracy starts to get worse, but training loss/accuracy is still improving, it is a likely result of overfitting

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

- Prepruning - ending the decision tree algorithm early when overfitting is noticed

- Post-pruning - completing the full decision tree then removing branches that are causing overfitting

Name: ____________________________

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).

i. This is explained by the fact that smokers take more breaks
ii. This is explained by the fact that Nicotine inhibits the CTS-inducing gene
iii. ~~This is explained by the fact~~ that non-smokers also have poorer postures
iv. All of these explanations are speculative
v. ~~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. America Must Win the race for A.I. Ethics (Fortune.com, 2/15/2022)

Link: https://fortune.com/2022/02/15/america-must-win-the-race-for-a-i-ethics-tech-artificial-intelligence-politics-biden-dod-will-griffin

Comment:

The article was written in response to the signing of the National Defense Authorization Act for Fiscal Year 2022, which contained 2 acts pertaining to AI, the Artificial Intelligence Capabilities and Transparency Act (AICT) & the Artificial Intelligence for the Military Act (AIM). I did not even know those two acts existed so that alone was news to me. The acts serve to guide the Department of Defense as they seek to compete with China in A.I. advancements, & put decision making power in this area with the Director of the National Science Foundation's Artificial Intelligence Research Institutions. The author in response suggested 3 ideas for maintaining ethical practice under these new acts; creating an AI use case archive, combining existing A.I. vetting frameworks, & developing a public communications strategy for AI related research & work.

I like the way the author presented these ideas in the article. Personally, I fall on the risk averse side of AI with regards to ethics, especially when involving the military. I feel concerned that someday warfare & the lives of people can be influenced by imperfect machine learning models. The author, however, doesn't fear monger or attack these developments. Instead he highlights alot of existing research, frameworks, & application examples to make the point that ethical use is already something being handled & studied.

I agree most strongly with his point of unification of AI vetting standards & procedures, not just in military but also private industry. It doesn't make sense to have AI ethical verification become a privitized, for-profit effort run by individual companies. If the research can be combined & applied through a unifying regulating body, I think that is the best way to ensure maximum impact. That said, given the U.S. stance on increasing government regulation, especially when it will impact private businesses, I could see that becoming a controversial issue in the future.

Overall, the article was current and gave interesting opinions on possible solutions while maintaining a hopeful & positive tone about the future of our government & AI/Machine Learning techologies