**Assignment 2: MLPS(adewan@andrew.cmu.edu)**

**Question 1: Conceptual: Logistic Regression**

**A close up of text on a white background

Description generated with very high confidence**

**A close up of text on a white background

Description generated with very high confidence**

**3.**

**A close up of text on a white background

Description generated with very high confidence**

**4.A close up of text on a white background

Description generated with high confidence**

**A close up of text on a whiteboard

Description generated with high confidence**

**A close up of text on a white background

Description generated with high confidence**

**Question 2:** **Conceptual: Decision Trees [25 points]**

**1.**

a) True: Yes, it is possible to achieve a 100% train accuracy if we memorize the train dataset completely in decision tree we can do this my creating tree paths for every dataset row which is worst approach.

b) False: No, a same node can be used in decision tree more than once in a different path or may be it can be used in left tree branch once and in the right branch later.

**2.a)**

**A close up of text on a white background

Description generated with very high confidence**

**A close up of text on a white background

Description generated with high confidence**

**2 b)**

**A close up of text on a white background

Description generated with high confidence**

**3.**

a)As she is considering splitting only on nodes where information gain is positive and not splitting on nodes which have 0 information gain.(as info gain can be either 0 or positive and cannot be negative). So she is not letting her tree grow when IG=0

The following dataset can be used.

|  |  |  |
| --- | --- | --- |
| X1 | X2 | y |
| F | T | T |
| T | T | T |
| F | T | F |
| T | T | F |
| F | F | T |
| T | F | T |
| F | F | F |
| T | F | F |

b) The flaw is in the termination condition: Terminating when IG=0 which is incorrect.

C) Correction to the flaw is that we need to consider splitting on nodes that have 0 info gain too if that is the best possible available node split.

A close up of text on a white background

Description generated with very high confidence

**4.**

i) 2(v-1) -1

ii) v(2(v-1) -1)

iii)(v-1) (2(v-1) -1)

5.

A close up of text on a white background

Description generated with high confidence

6.

A close up of a piece of paper

Description generated with high confidence

7.

A close up of text on a whiteboard

Description generated with high confidence

A close up of text on a whiteboard

Description generated with high confidence

A close up of text on a white background

Description generated with high confidence

**Question3: Model Evaluation and Model Selection(Conceptual)**

1. False:

The best value of hyperparameter is not the average of hyper parameter values for each fold.

But to evaluate the best possible values of hyperparameter we use techniques like cross validation and Grid search that exhaustively searches for best possible value of hyperparameter by plugging various values on a hyperparameter grid and then reports the value that produces best evaluation metric value (like accuracy score)

1. True
2. False:

No, train accuracy score can never give us a true picture of test accuracy score. A very low train accuracy score can still have a high-test accuracy and vic-a-versa.

1. True
2. True:

Yes, if the model is trained on the training data and it leads to high train accuracy if as the model gets trained on the dataset. Then cross validation accuracy will be significantly dropped as it reports the true accuracy of the model without any memorization.

1. False:

Cross validation accuracy basically measures the accuracy of the model by taking various folds of data but still it is the part of training set as a result it reports an unbiased and true train accuracy for the model. This cross-validation accuracy is in fact higher than test accuracy as test data has unobservable characteristics which the model has not been perfectly trained on which adversely affects test accuracy.

1. False:

Numerical hyperparameter are difficult to optimize but ranging over various values to calculate accuracy and keep on incrementing the hyper parameter in some step size and recording the accuracy we will find a turning point(sweet-spot) where the accuracy will bes the maximum and above which the model starts getting overfitting. This is the point of low variance and bias which we can say to be optimal sweet spot.

2.

Accuracy is not always a great measure to go about, there are certain flaws if we only evaluate a model on basis of accuracy as:

* it considers FP and FN as same which should not be considered as same as some time one has bigger impact than other.
* Secondly it do not consider any imbalance in classes in the test dataset and so may give a good train accuracy but perform poorly on test data

**Apart from accuracy matrix we can use AUC ROC:** In this, the data is split in train, validation and test data, and it do consideres imbalance in classes and considers difference between FP and FN,

**In the procedure**:

* the AUC scores are obtained for all validation sets across all the folds in a cross-validation procedure
* mean validation AUC score is then calculated this is used for giving the validation performance metrics (of AUC score).
* Similarly test AUC score is calculated on held out test data.

3.

She Is doing **wrong methodology**, she **is using test errors** and looking on the performance of test errors and plotting it, she is tuning her hyperparameter by increasing the depth of tree as per the results obtained from test error.

This is a wrong approach and can lead to **wrong and biased estimates**. Test error should be used not for model selection and model training purposes but only for model evaluation once and for all.

**STEPS SHE SHOULD FOLLOW:**

She should take different values of D at a larger space like D {5,10,15,20,25,30,35….100}

Now cross validate on all depths and report the results and plot it.

She might not find exact 78 as the right depth but she can surely decide on 75 and get little high-test error as compared to most optimal depth size of 78.

But this model will be more **robust** and **accurate** and **test data would not be touched in training time.**

Discussed ideas with:

Karunashree Saproo(ksaproo)

Sathwik Chenna Madhavuni(schennam)