

COMPSCI 762 QUIZ

Aiden Burgess, abur970, 600280511

1.

“For the duration of this quiz, I, Aiden Burgess, confirm that I will not discuss the content of the quiz with anyone else for the next 24 hours. I will not give any assistance to another student taking this quiz. I will not receive any assistance from any person or tutoring service. I will not post any information about this quiz on the Internet.”

2.

Scenario 1

- Pros: Validation error computed only once. Unbiased approximation of the test error. Training score is more accurate and closer to test score. Small optimization bias
- Cons: 1/2 data is not used to train the model. Only ran once so the accuracy may be randomly at 85%

Scenario 2

- Pros: Small optimization bias
- Cons: 1/2 data is not used to train the model. Can't tell if the 85% accuracy is reliable because there is no comparison to the other 9 models, this could be the best model just by random chance, and therefore may not perform well against real data.

Scenario 3

- Pros:
- Cons: 1/2 data is not used to train the model. Can't tell if the 85% accuracy is reliable because there is no comparison to the other 9999 models, this could be the best model just by random chance, and therefore may not perform well against real data. Validation error has been computed too many times. High optimization bias

Scenario 4

- Pros: Have averaged the accuracy to give a closer measure to real world performance. Small optimization bias
- Cons: 1/2 data is not used to train the model

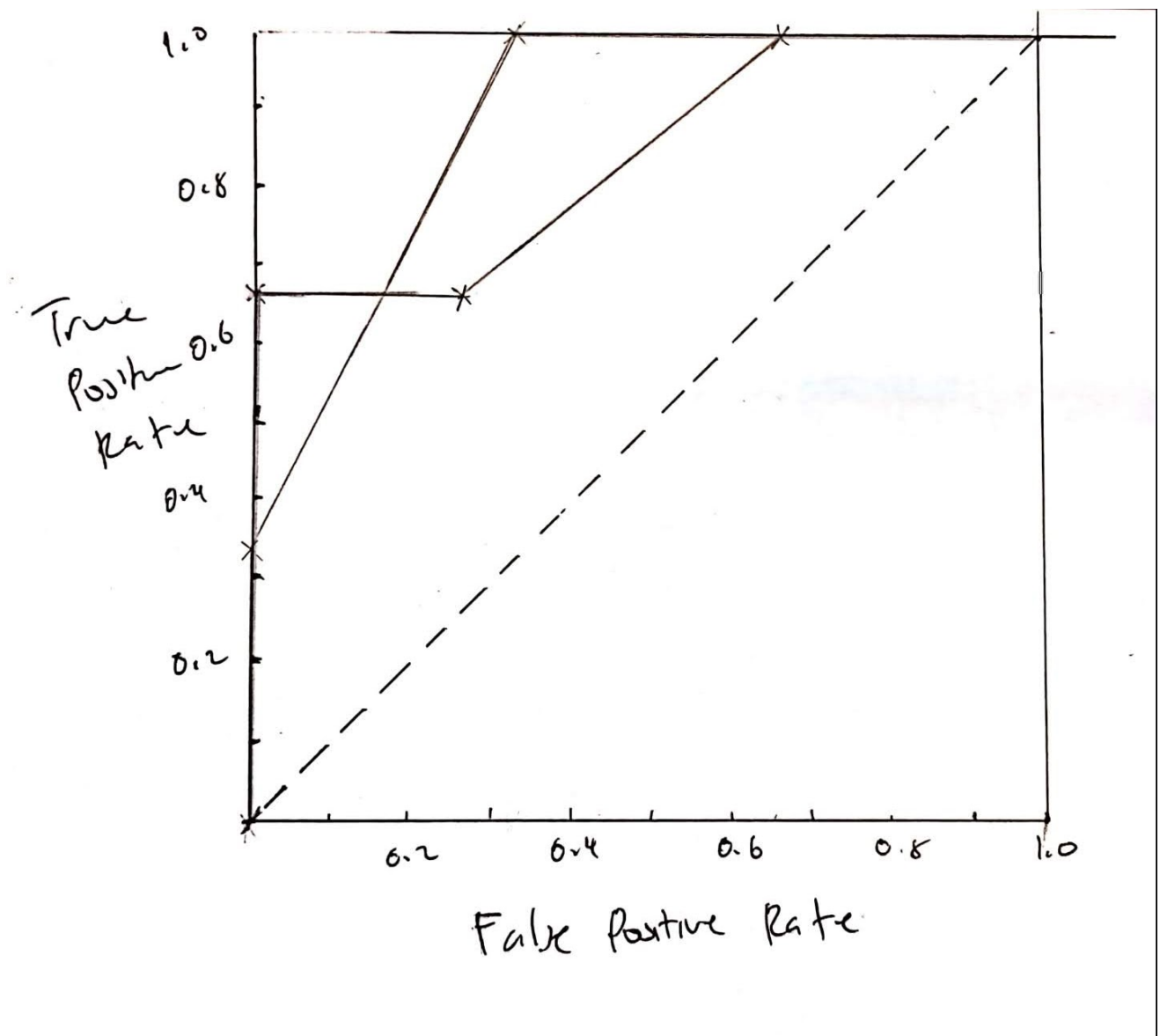
Will use scenario 4. Scenario 4 reduces confirmation bias as it only runs the model once. The score is also averaged across 10 models to reduce randomness of model score. This is under the assumption that the data is split randomly each run and there is a different random seed each time.

Scenario 2 and 3 have scores that may not be accurate at all because they could randomly be at 85%, and the choice of choosing what the score to report is unknown. Scenario 1 is the next best option, but only 1 model is trained which could be at 85% by random chance.

3.

It is appropriate if by chance the test set data is not representative of the overall data. It is not appropriate if the test set data is representative, as it is giving an accurate prediction. By repeating the test until a desirable score is achieved, the actual predictivity of the model is unknown. The highest scoring accuracy could be by chance.

4.



Classifier 1 is the best classifier as it has a greater area under the curve than classifier 2. This means it is a more accurate classifier.

A	B	C	D	E	F	G	H	I	
Class	CLASS1	THRESHOLDS	0	0.2	0.4	0.6	0.8	1	
P	0.83		1	1	1	1	1	0	
N	0.78		1	1	1	1	0	0	
P	0.66		1	1	1	1	0	0	
P	0.62		1	1	1	1	0	0	
N	0.48		1	1	1	0	0	0	
N	0.32		1	1	0	0	0	0	
TPR			1	1	1	1	0.33	0	
FPR			1	1	0.66	0.33	0	0	
Class	CLASS2	THRESHOLDS	0	0.2	0.4	0.6	0.8	1	
P	0.92		1	1	1	1	1	0	
N	0.62		1	1	1	1	0	0	
P	0.89		1	1	1	1	1	0	
P	0.52		1	1	1	0	0	0	
N	0.49		1	1	1	0	0	0	
N	0.38		1	1	0	0	0	0	
TPR			1	1	1	0.66	0.66	0	
FPR			1	1	0.66	0.33	0	0	

5.

Regularization is meant to fix overfitting of a neural network. It does this by making sure the weights for each node are small (around 0). By doing this, no single path or weight can dominate predictions, and weights are kept in linear area of regression. It moves the model from high variance towards higher bias. One type of regularization is dropout regularization. Each node has a possibility of not being included in the model when being trained for each batch. The probability of a node being kept in the model being trained is called the keep probability.

6.

Minimal depth of decision tree is 1, which is a decision stump. AKA one split.

11 data points are miss-classified. All 11 miss-classified points are from C1 being classified C2.

