# Quiz 3

This is a preview of the published version of the quiz

Started: Oct 6 at 12am

# Quiz Instructions

## Question 1 — 1 pts

You work at a bank and want to impress your boss, so you develop a binary classification model that predicts whether a customer will pay back their loan or not. You run your model and find out that the model has a **recall** of 80%. Determine which of the following options is the correct implication of using the classifier.

- ◯ Out of all the loans our bank gave to customers, 80% will pay them back.
- ◯ Out of all the loans our bank gave to customers, 20% will pay them back.
- ◯ We missed 20% of people that would have paid us back by rejecting them.
- ◯ We missed 80% of people that would have paid us back by rejecting them.

## Question 2 — 1 pts

Suppose you have two models A and B evaluated on the same test data of a classification task and you observe the following results:

# of examples misclassified by both models $N_{00}$ = 45

# of examples misclassified by A but not B $N_{01}$=25

# of examples misclassified by B but not A $N_{10}$=8

# of examples misclassified by neither A nor B $N_{11}$=150

Use the McNemar's test and a significance threshold of 0.05 to determine which one of the following statements is correct.

- ◯ The test statistic is around 8, model A is significantly better than model B.
- ◯ The test statistic is around 8, model B is significantly better than model A.
- ◯ The test statistic is around 5, model A is significantly better than model B.
- ◯ The test statistic is around 5, model B is significantly better than model A.

## Question 3

1 pts

Determine the recall, precision, and accuracy (rounded to the nearest hundredth) of a binary classifier given that its performance is provided in the following confusion matrix:

| | | **Actual Label** | |
|---|---|---|---|
| | | **True** | **False** |
| **Predicted Label** | **True** | 100 | 10 |
| | **False** | 20 | 110 |

- ◯ Recall=0.91, Precision=0.83, F1=0.87
- ◯ Recall=0.83, Precision=0.91, F1=0.87
- ◯ Recall=0.83, Precision=0.88, F1=0.85
- ◯ Recall = 0.88, Precision = 0.91, F1=0.89

## Question 4

1 pts

Select all strategies below that can help prevent or reduce overfitting in decision trees:

- [ ] Restricting the depth of the decision tree.
- [ ] Pruning the decision tree based on a validation set accuracy.
- [ ] Use more features to represent each examples.
- [ ] Use less features to represent each examples.

## Question 5

**1 pts**

We run the ID3 algorithm for learning decision trees on 800 instances <(A, B, C, D), y> where y is a binary label and A, B, C, D are binary attributes. It so happens that :

(i) 300 of the data points have A=0, and they split evenly between positive (y=1) and negative (y=0) examples. But when A=1, all the examples are negative.

(ii) 500 of the data points have B=0, but only 400 of them are negative (y=0) and the rest are positive (y=1) examples. Similarly, when B=1, only 50 of them are positive, and the rest are negative.

(iii) C and D take only the value 1, in all the examples.

Determine which of the following statements is correct:

- ◯ 18.75% of the examples are positive and A is chosen to be the root node.
- ◯ 18.75% of the examples are positive and B is chosen to be the root node.
- ◯ 25% of the examples are positive and there is a tie between C and D on who is the root node.

Quiz saved at 12:00am Submit Quiz