

# QUIZ:- 10

Started on	Dienstag, 18 Januar 2022, 2:18
State	Finished
Completed on	Dienstag, 18 Januar 2022, 4:01
Time taken	1 hour 42 mins
Grade	3.33 out of 4.00 (83%)

Question **1**  
Complete  
Mark 0.50 out of 0.50  
🚩 Flag question

Which of the following statements about classification metrics is true?

Select one or more:

- ☒ A classifier that always predicts the most frequent class can achieve high accuracies when the class distribution is imbalanced.
- ☒ For a binary data set in which the two classes have the same frequency, a random classifier (a classifier that flips a coin for its prediction) would achieve around 50% accuracy.
- ☐ A classifier can always achieve perfect precision on a task if its threshold is set such that the recall is close to one on a test data set.

Question **2**  
Complete  
Mark 0.50 out of 0.50  
🚩 Flag question

How many splits/folds of your data set are needed at least for training and evaluating a machine learning model, with a fixed set of hyperparameters (that means without model selection/hyperparameter optimization)?

Select one or more:

- ☐ 3
- ☒ 2
- ☐ 1

Question **3**  
Complete  
Mark 1.00 out of 1.00  
🚩 Flag question

When performing cross-validation for model selection and model evaluation, different strategies are possible. Which of the following statements is true?

Select one or more:

- ☒ When you have to perform validation (for instance for hyperparameter optimization) then splitting your data into a training set, a validation set (for model selection) and test set for evaluating your model will give you a better estimate of the generalization error than splitting your data into training and test only as you do not allow your model to overfit its hyperparameters to the test set.
- ☒ The (computationally) most expensive estimate of the generalization performance is obtained using (nested) k-fold cross-validation.
- ☐ When model training takes very long, nested cross-validation can help to reduce the overall evaluation time.

Question **4**  
Complete  
Mark 1.33 out of 2.00  
🚩 Flag question

Let  $U$  denote the number of hyperparameter candidates, with  $U > 1$ , and let  $K$  refer to the number of cross-validation partitions/folds.

Which of the following statements are true?

Select one or more:

- ☐ Performing nested k-fold cross-validation with grid search for hyperparameter optimization will require to train the model just once.
- ☐ When performing cross-validation without hyperparameter optimization,  $U \cdot K$  models need to be trained.
- ☒ Performing nested k-fold cross-validation with grid search for hyperparameter optimization and evaluation will require to train the model  $K(((K-1)U)+1)$  times.
- ☒ When nested crossvalidation takes too long, one can split the data in three folds, one for training, one for model selection / hyperparameter optimization (the validation fold), and one fold for evaluation (the test fold) to both optimize hyperparameters and evaluate the generalization performance of the best model.
- ☒ When performing grid search for the model selection / hyperparameter optimization, the model needs to be trained at least  $K$  times.