

<b>Lab assignment evaluation: Classification</b>
--

Bachelor in Mobile and Space Communications Engineering

Bachelor in Telematics Engineering

Bachelor in Sound and Image Engineering

Bachelor in Telecommunication Technologies Engineering

**Name and surname:** \_\_\_\_\_

1. Which technique has been employed in the assignment to determine the optimal value of  $k$  in the  $k$ -Nearest Neighbors ( $k$ -NN) algorithm?

☒ **Grid search.**

☐ Using a validation set.

☐ Linear Discriminant Analysis (LDA).

☐ None of the above answers is correct.

2. What does the attribute `best_params_` included in the class `GridSearchCV` return?

☐ The estimator which gave highest score on the left out data.

☐ The mean cross-validated score of the `best_estimator_`.

☐ The index which corresponds to the best candidate parameter setting.

☒ **The parameter setting that gave the best results on the hold out data.**

3. Point out which statement, according to the performance of a binary classification algorithm, is correct.

☐ If the value of AUC is 0, there is a classification threshold for which the accuracy rate is 1.

☐ If the value of AUC is 1, there is a classification threshold for which the error rate is 1.

☒ **If the value of AUC is 1, there is a classification threshold for which the error rate is 0.**

☐ If the value of AUC is 0, there is a classification threshold for which the error rate is 0.

4. Have a look at the following code snippets:

```
from sklearn.metrics import roc_curve
FPR, TPR, Thresholds = roc_curve(y_test, scores)
```

What do `scores` represent?

- ☐ The probability estimates of the negative class.
  - ☐ The predicted class labels for samples in the test subset.
  - ☒ **The probability estimates of the positive class.**
  - ☐ The mean accuracy on the given test data and labels.
5. Point out which statement, according to the confusion matrix, is false.
- ☐ The sum of the diagonal elements indicates how many samples have been correctly classified.
  - ☐ It can be approximately computed both the probability of missing and the probability of false alarm from the elements of the matrix.
  - ☐ It can be approximately computed both the accuracy and error rates from the elements of the matrix.
  - ☒ **It is a symmetric matrix.**
6. Point out which statement, according to the ROC curve calculated in the assignment, is false.
- ☐ The ROC curve is increasing.
  - ☐ It passes through the point (0, 0).
  - ☐ It passes through the point (1, 1).
  - ☒ **It passes through the point (0, 1).**
7. Point out which statement, according to the experiments carried out in the assignment, with the class `RandomForestClassifier` in `sklearn`, is correct.
- ☐ It is necessary to standardize the data aiming at achieving good performance.
  - ☐ The hyperparameters which have been fine-tuning are **penalty** and **C**.
  - ☐ It is composed of a set of classifiers based on the  $k$ -NN algorithm (class `KNeighborsClassifier`).
  - ☒ **None of the above answers is correct.**
8. Have a look at the following code snippets:

```
from sklearn.linear_model import LogisticRegression
LogisticRegression().fit(X_train_s, y_train)
```

What does the method `fit` in the class `LogisticRegression` perform?

- ☐ It returns the mean accuracy on the given training data and labels.
- ☐ It predicts class labels for samples in `X_train_s`.
- ☐ It obtains the value of **C** (inverse of regularization strength).

- ✓ **It computes the coefficients of the model according to the given training data.**

9. Which of the following code snippets demonstrates the way used in the assignment to standardize the feature values for the training subset?

- ✓ `from sklearn.preprocessing import StandardScaler  
X_train_scaled = StandardScaler().fit_transform(X_train)`
- ☐ `from sklearn.preprocessing import StandardScaler  
scaler = StandardScaler()  
X_train_scaled = scaler.fit(X_train)`
- ☐ `from sklearn.preprocessing import Normalizer  
scaler = Normalizer()  
X_train_scaled = scaler.fit_transform(X_train)`
- ☐ None of the above answers is correct.

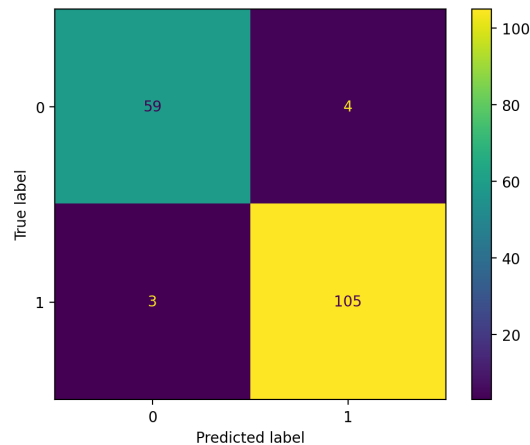
10. Have a look at the following code snippets:

```
v_ntrees = [10, 20, 50, 100, 125, 150]  
depth = [5,10,15,20]  
parameters_rf = {'n_estimators': v_ntrees, 'max_depth': depth}  
rf = RandomForestClassifier()  
grid_rf = GridSearchCV(rf, parameters_rf, cv=5)  
grid_rf.fit(X_train, y_train)
```

How many classification models are created and fitted when the method `fit` is runned?

- ☐ 10 models aiming at obtaining the optimal values of the hyperparameters + 1 (final) model using the aforementioned values.
- ☐ 24 models aiming at obtaining the optimal values of the hyperparameters + 1 (final) model using the aforementioned values.
- ✓ **120 models aiming at obtaining the optimal values of the hyperparameters + 1 (final) model using the aforementioned values.**
- ☐ 50 models aiming at obtaining the optimal values of the hyperparameters + 1 (final) model using the aforementioned values.
11. What is the significance of an AUC value of 0.5 in a Receiver Operating Characteristic (ROC) curve?
12. What is the impact of choosing a smaller value for  $k$  in the  $k$ -Nearest Neighbors ( $k$ -NN) algorithm?

13. It is displayed the confusion matrix for a binary classification problem. Having a look at the aforementioned matrix, calculate both the probability of missing and the probability of false alarm. Indicate how you do it.



14. Have a look at the following code snippet in which `model` represents classification model in `sklearn`.

```
fpr,tpr,thresholds = roc_curve(y_test,model.predict_proba(X_test_s)[: ,1])
```

This method returns the following output vectors:

```
fpr: [0 0.01 0.05 0.1 0.15 0.2 0.95 1]  
tpr: [0 0.1 0.8 0.85 0.9 0.95 0.98 1]  
thresholds: [0.95 0.9 0.7 0.6 0.5 0.3 0.2 0.1]
```

If you want to obtain a probability of detection equal or greater than 0.95, which is the minimum value of false alarm rate that you will obtain? For this working point, what is the classification threshold?

15. Which hyperparameters have been fine-tuning for the random forest classifier in this assignment? Explain, briefly, what each of them represents.