Pattern Recognition
and Applications Lab

Performance Evaluation

Machine Learning – Course Laboratory

Battista Biggio

battista.biggio@diee.unica.it

Luca Didaci

didaci@diee.unica.it

Dept. Of Electrical and Electronic Engineering

University of Cagliari, Italy

University
of Cagliari, Italy

Department of
Electrical and Electronic
Engineering

Ex. 1: Testing performance on unseen data

- To assess classifier performance, one should estimate the classification error on never-before-seen data
- The training data should not be used to this end, as it provides an optimistic estimate of the real performance!

Ex. 1: Testing performance on unseen data

1. Sampling a training and a testing set (from the same underlying distribution), e.g., splitting data x,y at random in x_tr, y_tr, x_ts, y_ts
   - `from sklearn.model_selection import train_test_split`

2. Normalizing training and test data (using parameters estimated on training data!)
   - `from sklearn.preprocessing import MinMaxScaler`

3. **Estimating classifier parameters on training data (next time!)**

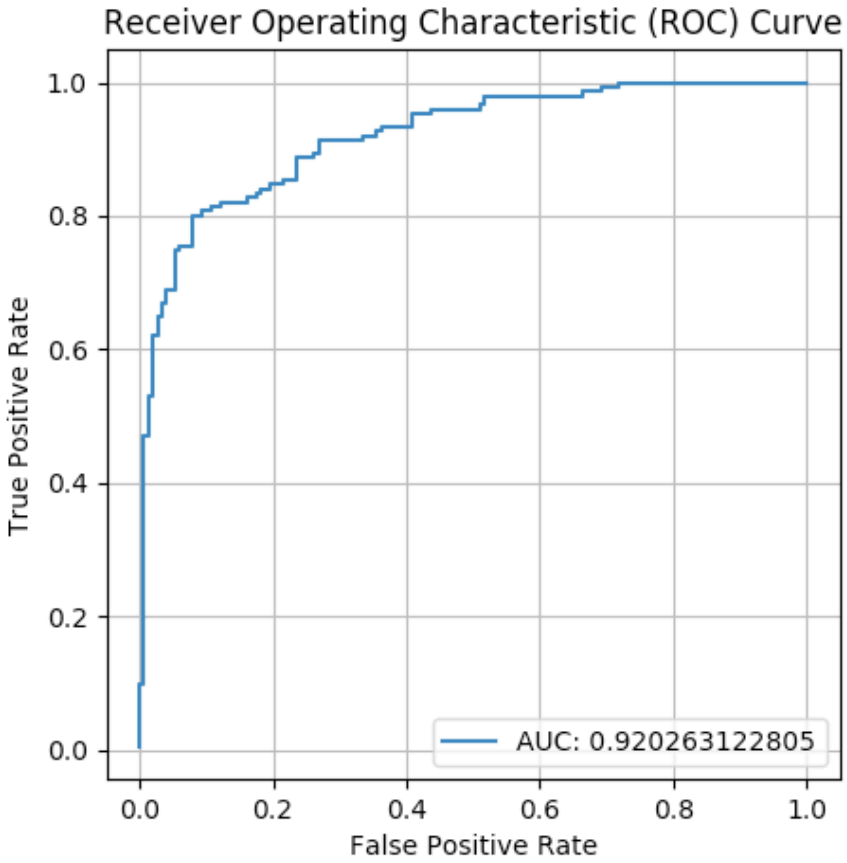
4. Fitting the classifier on training data
   - `clf.fit(x_tr, y_tr)`

5. Predicting the class labels of testing data - `clf.predict(x_ts)`

6. Evaluating accuracy or classification error

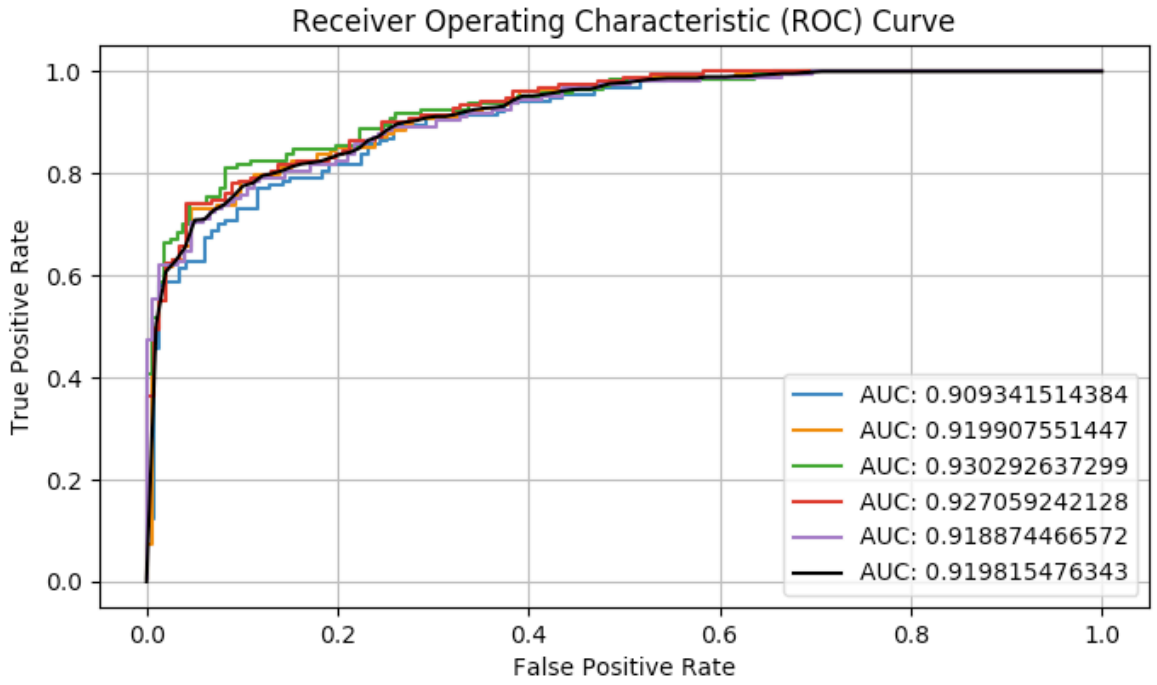
Ex. 1: Receiver Operating Characteristic (ROC) Curve

- For two-class classification problems, a common measure of performance is the so-called ROC Curve
- AUC corresponds to the Area Under the ROC Curve



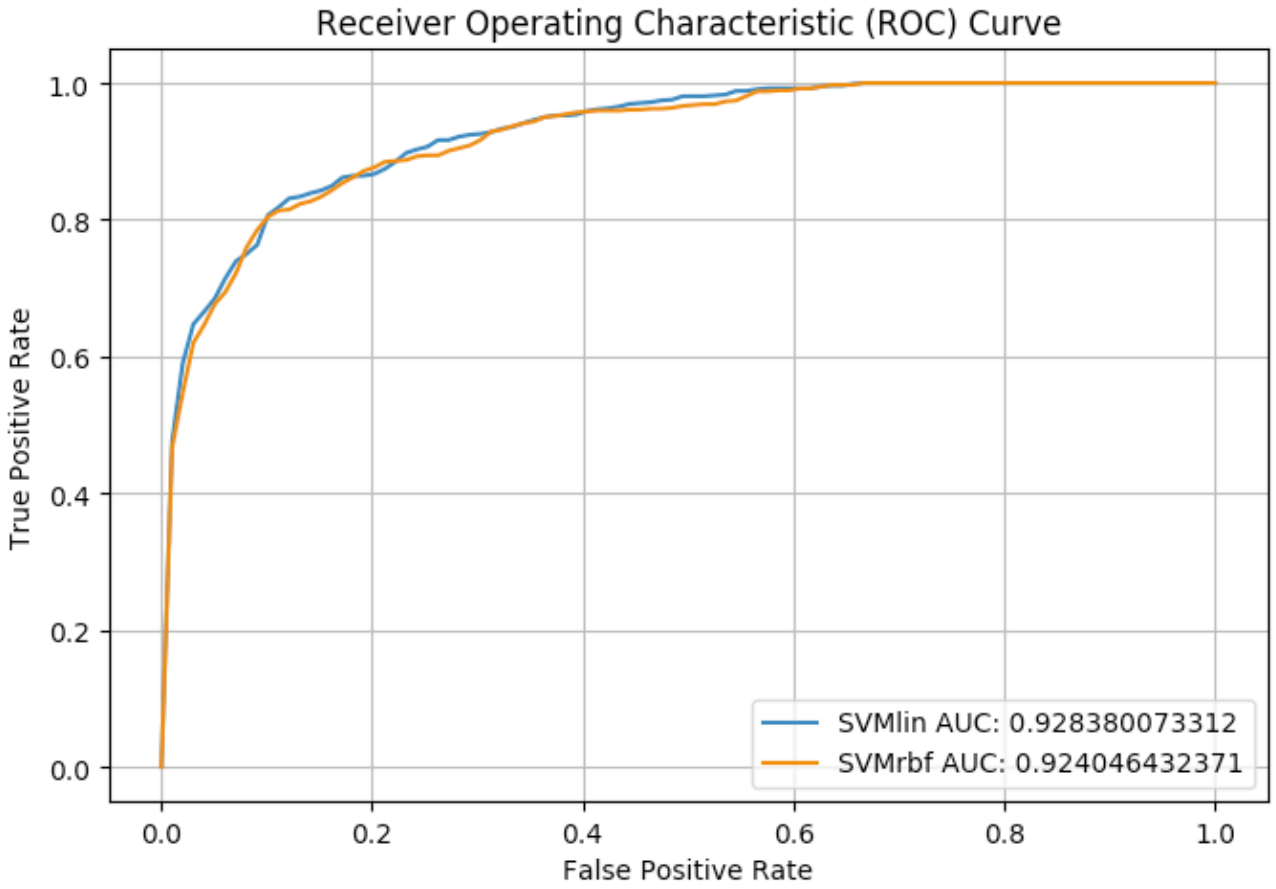
Ex. 2: Repetitions and averaging

- Evaluating performance on a single training-testing split does not give us any result that is ***statistically significant***
- To this end, one should evaluate perfomances on several independent training-testing splits, and average performances
- Average ROC curve (black line) on 5 repetitions



Ex. 3: Comparing different classifiers

- Average ROC curves of different classifiers



Lessons learned

- Estimation of classifier performance on unseen data
- Average ROC curves / classification accuracy

Student challenges:

- Compare different classifiers on (some) two-class datasets (of your choice), and report for each the average ROC curves,
  - using the same data splits for the different classifiers in the same repetition!

- Compare different classifiers on multi-class datasets, including IRIS and MNIST Handwritten Digits, in terms of average classification accuracies

Please e-mail us if you are able to solve any of them!