
Lab 7: Classification (Multilayer Perceptron)

Exercise 1:

You are an employee on a golf course, under the direct supervision of the owner, who every days makes the decision of opening or closing the course based on several circumstances: the weather (Sunny, Overcast, or Rainy), the Temperature (Hot, Mild, or Cool), the Humidity (High or Normal), and the Wind (Weak or Strong).

You have been working at this course for 14 days and have been observing the behaviour of the owner in making the open/close decision. You have collected the following data:

Day	Weather	Temp.	Humidity	Wind	Decision
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

This morning, the owner has called in sick and is asking you to make the decision of opening or closing the course based on what you have learned so far. For this, you think the best idea is to build a decision tree based on your data. You proceed to choose which of the attributes should be used to make the first split in your tree by computing the Gini index for each of them.

Exercise 1.1: Compute the Gini index associated to each attribute

Exercise 1.2: Based on your calculations, which attribute should be used to make the first split on your data? How would you proceed now to complete your decision tree?

Exercise 1.3: Use **scikit-learn** to train a decision tree on the above data. You now collect some measurements, and realize that today is sunny and cool, with a high humidity high and there is strong wind. Should you open the course?

Exercise 2:

You work for an online shopping business that wants to forecast user behavior and predict whether they will be making purchases or not based on previous browsing habits. Most visitors to our online shopping website, don't end up going through with a purchase during a web browsing session. It might be useful, though, for our shopping website to be able to predict whether a user intends to make a purchase or not: perhaps displaying different content to the user, like showing the user a discount offer if the website believes the user isn't planning to complete the purchase. How could our website determine a user's purchasing intent?

Exercise 2.1: Load the training data, divide it into a train and a test set, and train a **scikit-learn** decision tree on the training set. Evaluate its accuracy both in the training and in the test set. Do you see underfitting or overfitting?

Exercise 2.2: In order to improve the performance of your classifier, let us limit the maximum depth of the tree to 5 and repeat the training. What is the impact of reducing the depth on the performance of your classifier in the training, and in the test set?

Exercise 3: Implement and compare the following classifiers on the same dataset (shopping dataset):

- Naive Bayse Classifier
- MLP
- SVM
- Decision tree