PHY499: Introduction to Python for Scientists

Homework Assignment 8 (2 Dec, 2016)

(NOTE: This assignment will be graded!)

Send your code to `michael.mommert@nau.edu`.

1 Machine Learning Accuracy

The examples discussed in the lecture show that classifications derived with ML methods can be ambiguous and may depend on the choice of the classification method.

Of course, reality is even more complicated: the researcher measuring the petal/sepal lengths/widths that are listed in the array `new` did not account for measuring uncertainties. Assume that each number listed in array `new` is subject to uncertainties of the order of 0.2 cm that are distributed in a Gaussian way. What is the average probability of each of the 5 elements in `new` to belong to either class?

Task: Print a table with the average probabilities and corresponding standard deviations for each element in `new` to belong to either class using the Nearest Neighbor classifier (k=1 and k=10) and a linear SVM model in a Monte Carlo approach.

Hints: Use the following approach for each classfier method:

- train your classifier using the training sample (you can assume that the uncertainties on the training sample are negligible)
- vary your data sample according to the noted uncertainties in a Gaussian way (use `np.random.normal` to vary the data where option `loc` is the mean [the quoted length or width] and `scale` is the 1σ uncertainty [0.2 cm]); use your trained classifier on this modified data sample and derive the classification probabilities; repeat this 1,000 times (this is the Monte Carlo component)
- derive the average and the standard deviation for each element in `new` and for each class (subspecies) over all the Monte Carlo runs; print these numbers in a human-readable way on the screen
- repeat the aforementioned approach for the k=1 NN method, k=10 NN, and the linear SVM model

If you want this assignment to count into your final grade, please submit it to `michael.mommert@nau.edu` before 8 Dec, 23:59!