



Data Application Lab

Scikit-learn Machine Learning

1. When should you use classification over regression?

Classification produces discrete values and dataset to strict categories, while regression gives you continuous results that allow you to better distinguish differences between individual points.

You would use classification over regression if you wanted your results to reflect the belongingness of data points in your dataset to certain explicit categories (e.g. If you wanted to know whether a name was male or female rather than just how correlated they were with male and female names.)

2. How can you choose a classifier based on training set size?

If training set is small, high bias / low variance models (e.g. Naive Bayes) tend to perform better because they are less likely to be overfit.

If training set is large, low bias / high variance models (e.g. Logistic Regression) tend to perform better because they can reflect more complex relationships.

3. Answer of Notebook:

<https://drive.google.com/file/d/0B-TSrmZPrApfLVhfcVNEejRmTTg/view?usp=sharing>