Classification Empirical Study

Introduction. We introduced you to classification on Day 2. There are many ways to perform classification, you have so far seen only logistic regression in detail. In this project, you will become more familiar with two other very effective, but different, classification methods: random forests and K-nearest neighbours (not to be confused with K-means clustering, an unsupervised method).

Project description. In this project, you will learn about random forests and K-nearest neighbours. You will then apply these methods, out of the box from scikit-learn, to the ImageNet object classification data set. The main assignment is for you to make an empirical study on the performance, strengths, and weaknesses of these methods (including logistic regression) using a variety of test metrics that you will implement. The key skills you will develop will be to apply and evaluate among a choice of several different methods for a problem.

Dataset. The project will use the ImageNet object classification data set, consisting of images with an associated category for each image.