

Report for Sheet 03

Lab Course Machine Learning and Data Analysis

Mario Tambos

July 12, 2017

Implementation comments

The code was structured mostly in functions following an imperative paradigm, except for the `krr` and `SupervisedMethod` classes, which were required by the assignment's statement.

One function was declared for each assignment in Part 2.

Beyond `matplotlib`, `numpy` and `scipy`, `seaborn` was used to improve the plots aesthetics in Part 2.

All tasks were completed, except for the final point of Assignment 4, and all tests passed.

Assignment 3

Figure 1 shows the ROC curves for different sample sizes. The false positive (fpr_n) and true positive (tpr_n) rates for a sample of size n were calculated analytically as follows:

$$\begin{aligned}fpr_n &= (1 - \Phi(x_0; \mu = 0, \sigma^2 = 1)) \\tpr_n &= (1 - \Phi(x_0; \mu = 2, \sigma^2 = 1))\end{aligned}$$

where $x_0 \in [-5, 5]$ is the threshold (bias) term.

As seen in the plot, the empirical ROC curve converges to the analytical ROC curve asymptotically.

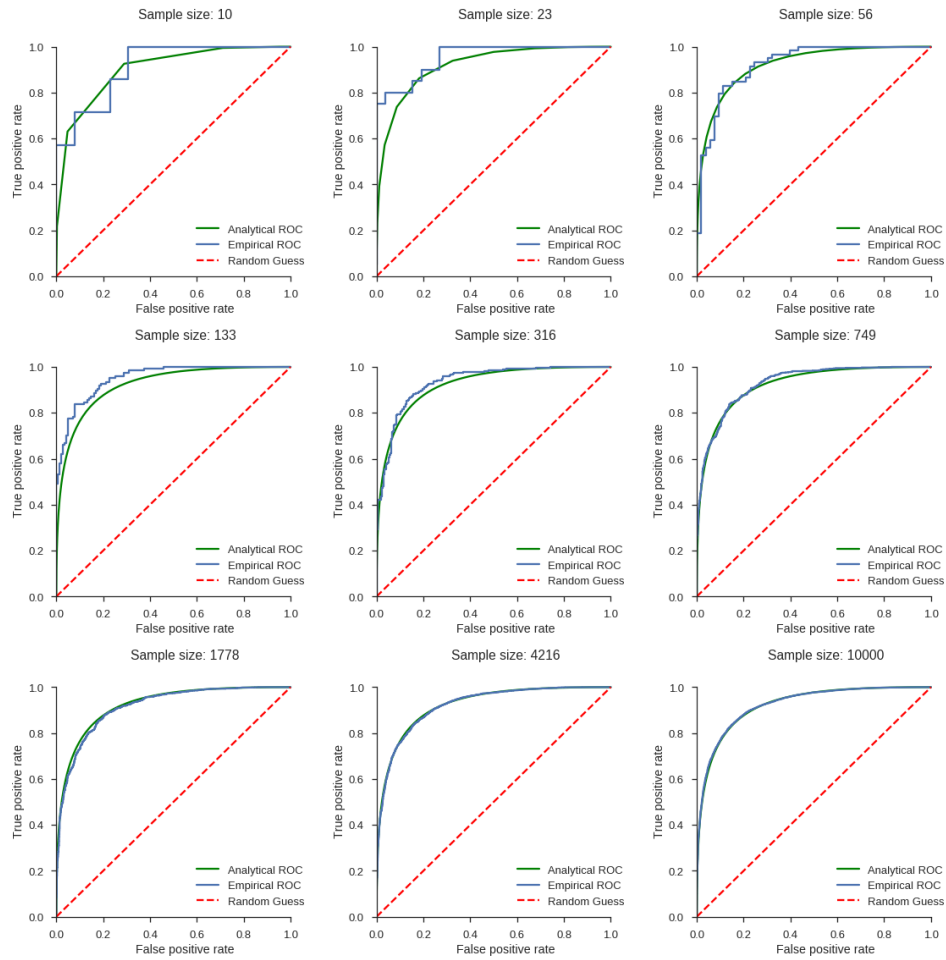


Figure 1: Analytical vs empirical ROC curves for different sample sizes.

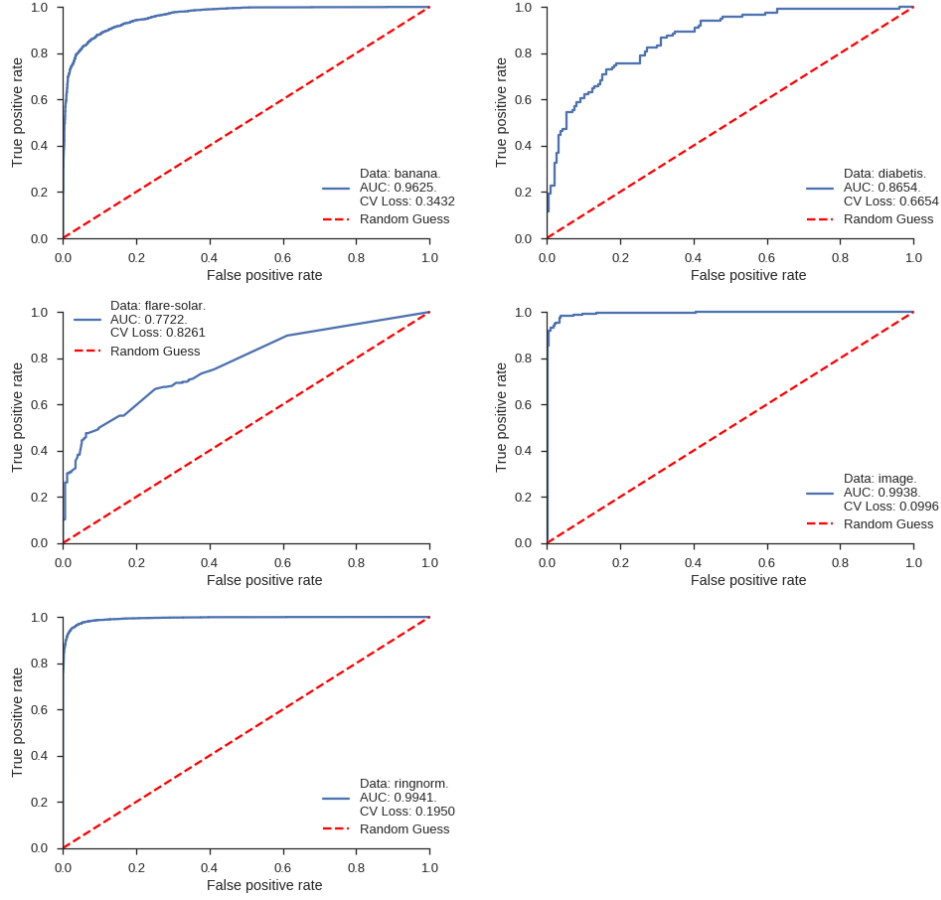


Figure 2: ROC curves for KRR resulting from variation of the bias term, with their respective AUC and cross-validation loss values.

Assignment 4

Figure 2 shows the ROC curves for the five datasets, with their corresponding AUC scores and cross validation losses.

From the experiments, it would seem that the AUC score has an inverse relationship to the cross validation loss; i.e., larger AUC scores correspond to lower cross validation losses.