Proef/oefen tentamen Oktober 2019, vragen

Machine Learning (Erasmus Universiteit Rotterdam)

Machine Learning
FEM31002 **Sample Exam 1-4 Weeks** September 19, 2019

1)(Week 1&2) Answer the following questions for K-nearest neighbors (K-NN) algorithm.

a. Explain how the selection of K for a K-NN classifier affect the model's flexibility, variance and bias.

b. Assume that your data can be classified with a linear Bayes decision boundary. Which value of K does give better results? Why?

2)(Week 2) Assume that you are given an assignment which asks you to perform nested cross-validation with repeated resampling on a data set. Give the pseudo-code of your solution.

3)(Week 3) Derive the Lasso regularization from the Bayesian point of view. Recall that

$$Laplace(\mu, b) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}}$$

4)(Week 4) Consider the likelihood function $\ell(\beta_0, \ldots, \beta_p)$ for logistic regression as we have discussed in the class.

a. Show that
$$\ell(\beta_0, \ldots, \beta_p) = \Pi_{i=1}^{n} p(x_i)^{y_i} (1 - p(x_i))^{1-y_i}.$$

b. Using part a, show that
$$\frac{\partial \log(\ell(\beta_0, \ldots, \beta_p))}{\partial \beta_j} = \sum_{i=1}^{n} (y_i - p(x_i)) x_{ij}.$$