Sven Hauns, Anup Kumar, Simon Bray
Machine Learning in Life Sciences
May 10, 2022

Excercise2: Support Vector Machines

Question 1. Download the dataset ELAVL1-PARCLIP and preprocess it, by featurizing the data using 3-kmers. Split the created dataset into training and test datasets (Hint: You can reuse your code from Ex.1 here).

Question 2. Build a classifier using:

(a) a soft-margin SVM
(b) a hard-margin SVM

Try out different kernel functions! Choose a suitable evaluation metric. How do you explain the differences? (Hint: make use of the sklearn library)

Question 3. Perform hyper-parameter optimization using grid search (on the Hyperparameters C and kernel of sklearn) Which values perform best?

(a) create a bar plot showing the distribution of a metric for one parameter.

Question 4. Explain some advantages and drawbacks of grid search. When would it be advantageous to use random search instead?