Algorithmic Data Analysis

Esther Galbrun
Autumn 2021
Suppose that a classification training algorithm requires \( O(n^r) \) time for training on a data set of size \( n \). Here \( r \) is assumed to be larger than 1. Consider a data set with a perfectly even distribution across \( k \) different classes.

As an example case, consider \( r = 3 \), \( n = 1000 \) and \( k = 5 \).

Compare the running time of the one-against-rest approach with that of the one-against-one approach.
Q1.2: Unbalance

Compare the characteristics of resampling vs. reweighting when dealing with unbalanced datasets.

efficiency
flexibility
out of the box
randomness
Boosting is a \{ \text{data-centered} \} \text{ensemble learning approach} \{ \text{model-centered} \} \text{aiming to reduce} \{ \text{bias} \} \text{\{variance\}.}
Q1.4: Bag and boost

Would you rather use a *linear SVM* or a *kernel SVM*?

i) To create the ensemble components in *bagging*

ii) To create the ensemble components in *boosting*