
FEATURE EXTRACTOR: CHI-SQUARED

- selecting subset of relevant features (genes) for use in model construction
 - simplification of model = easier to interpret
 - shorter run time
 - reduce overfitting
 - how likely it is that any observed difference between the sets arose by chance
 - calculate chi2 statistic for each sample, selecting top k features
 - $k = 50, 75, 100$
-

K-NEAREST NEIGHBORS (KNN)

- sample is classified by majority vote of its k nearest neighbors

