

CS 307 Lab - Week 1

1 Dataset

We have attached a csv file of a dataset for classifying Turkish raisins into types [here](#).

2 Questions In the Lab Session

Q1. Make a leave-one-out cross-validated estimate of the accuracy of a nearest neighbor classifier on this dataset, with no normalization of features.

Q2. Make a leave-one-out cross-validated estimate of the accuracy of a nearest neighbor classifier on this dataset, but now normalizing each feature to have mean 0 and standard deviation 1.

Q3. What are the standard deviations of each estimate?

Q4. Can you improve this classifier by using more than one nearest neighbor? How do you know that your improved classifier is better or worse?