

DATA 311 - Lecture 19: Supervised Learning

Classification, Regression

K nearest neighbors

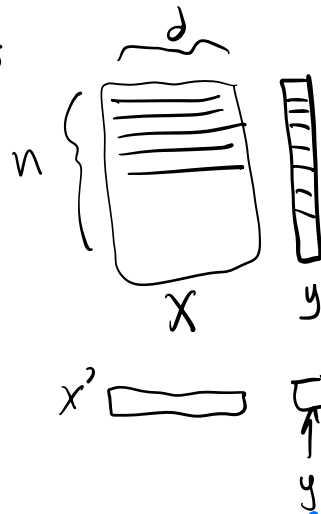
Hyperparameters and data splits

Dataset: $X : (n, d)$

$\uparrow$ # datapoints

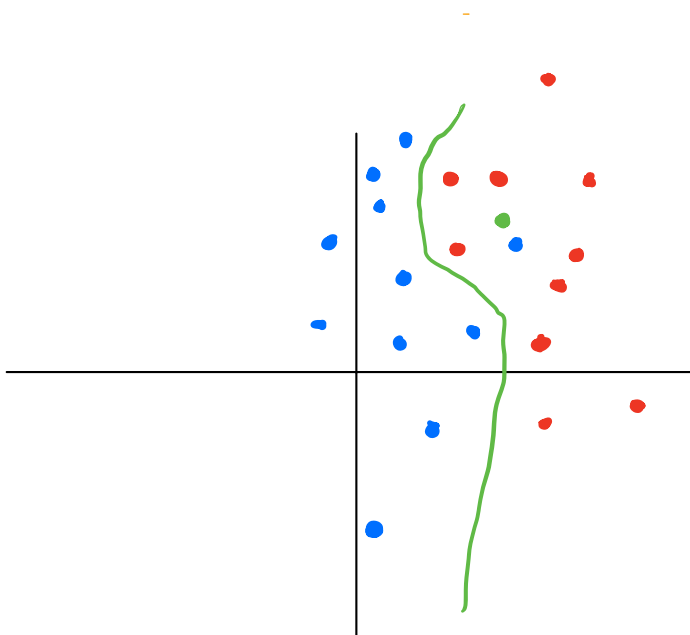
$\uparrow$ feat. dimensions

Labels: y

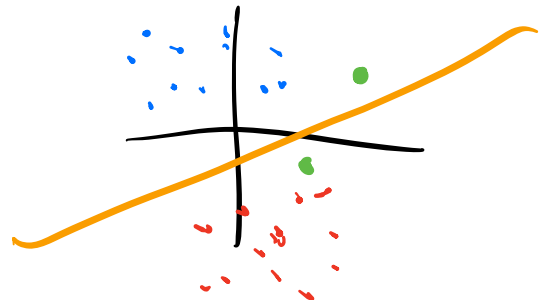


- Todo:
- Supervised
 - clas. vs reg.
 - NN, kNN
 - clas. + reg.
 - hyperparameter k
 - train/val split, tune k
 - (-) classification accuracy

If y is **categorical**, this is **classification**
If y is **numerical**, this is **regression**



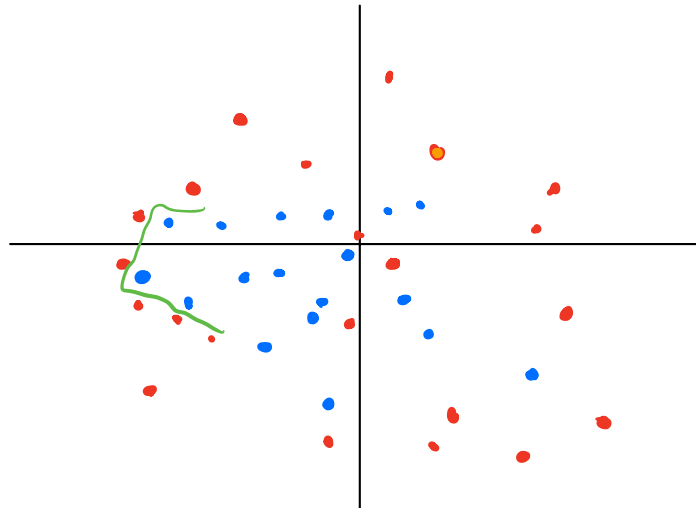
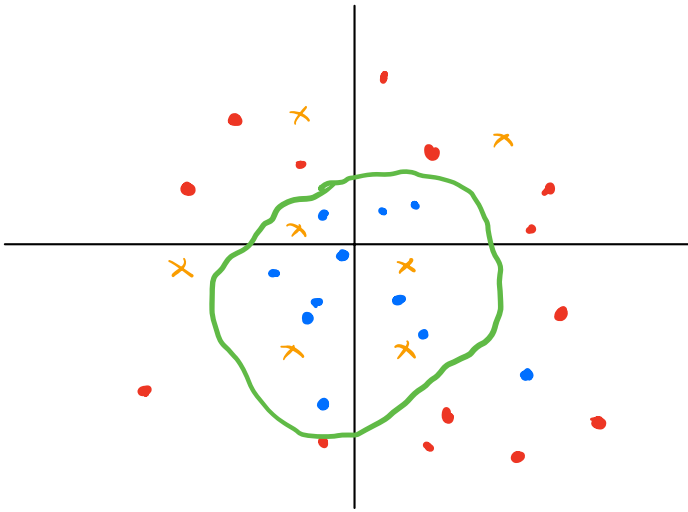
Oatmeal:
recipe $\rightarrow$ good / bad



kNN:

k is a hyperparameter

evaluate on different k
using validation set



training set

validation set

