K-nearest Neighbours

- handle both binary & Multiclass
- No assumptions about parametric form of the decision boundary

KNN has no training phase

Normal Seenario:

- N objects
- each rep by set of attrib X_n
- And a label t_n

to classify X_{new} w/ KNN we first find the k number of training point that are closest to X_{new}

t_{new} is then decided based on the majority class among the n ~~neigh~~ neighbours

One issue with ~~the exam~~ knn is the example of tied class group, i.e. k = 8 neighbour class 1 = 4 C2 = 4

Solutions -
- use odd k
- weight neighbours, closer = more influential. e.g. endicbar

How to choose k?

- too small = influenced by noise
  - This creates islands a clusters which represent overfitting

- Too big =
  - upto a point >k has a regularizatrai affect ~~and~~ which reduces overfitting
  - Too far then we loose true fit of the data
  - data inbal w/ hegi k can completely shut out a class or just overeach

K-cross validation

the most popular method for choosing k is k-cross valid

The metric we look @ to decide is the proportion of classification mistakes made

10-fold cross valid & repeat 100 time

Plot error result for each k

Select the minima

With knn & 10-Fold, train on 9 segments & validate on the 1 to count the errors

repeate cycling through the holdout segments