SVM Cont.

MARGIN: Distance from separation line to class of points

Good Margins: Separation distance is maximized between all classes

RANDOM FOREST

- A model containing multiple decision trees
- Can be used for regression and classification problems
- It can handle missing data and maintain accuracy
- Won't overfit data to model
- It can handle a large dataset with high dimensionality

DISADVANTAGES

1. Not too good for regression, the range of values depends on the range in training data
- VERY little control over what the model does

APPLICATIONS

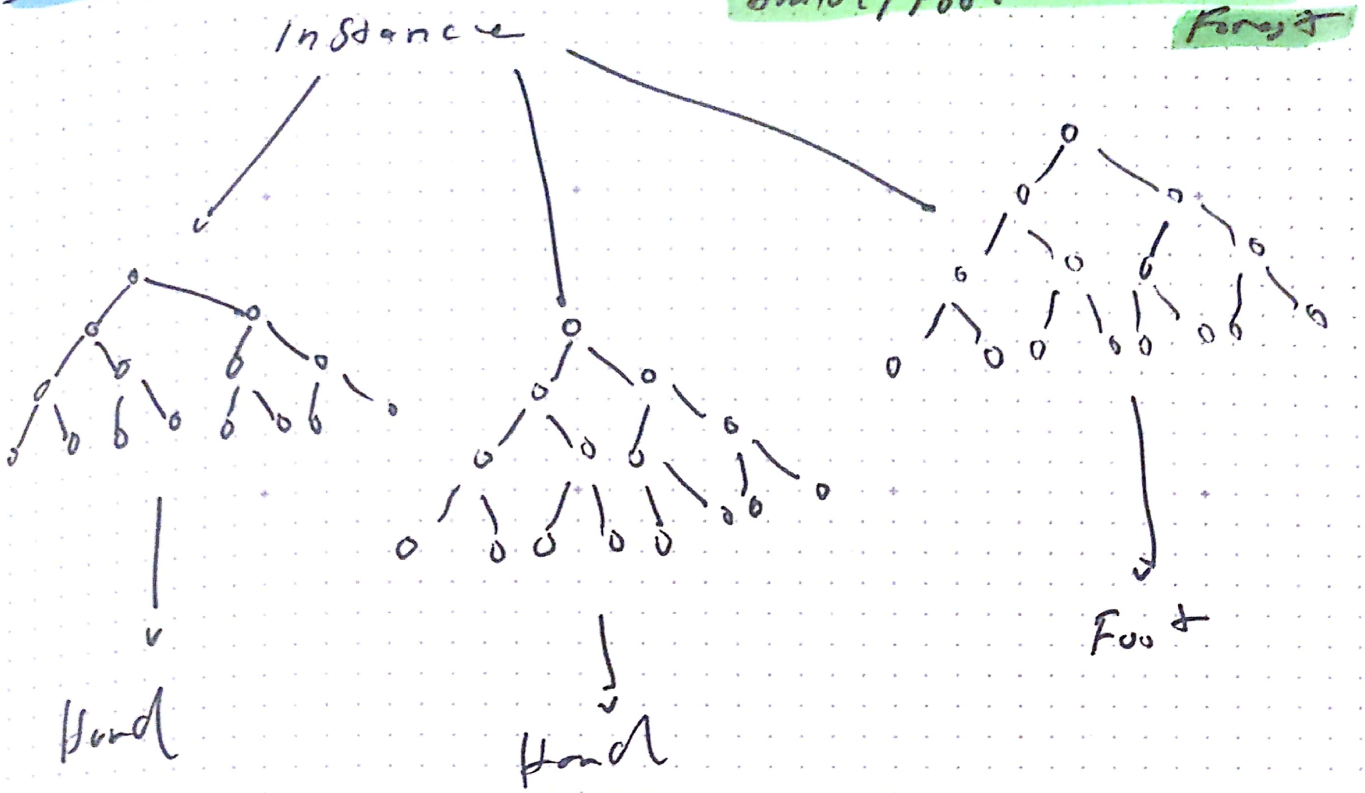
- Identify fraud
- Find stock behavior, cost/benefit of buying a share
- Customer product suggestions on consumer sites

How it works

Given a number of trees P, each tree decides on a value

RANDOM FOREST (cont.)

Example classifying Hand/Foot w/ Random Forest



Hand: 2/3

Foot: 1/3

Classification: Hand

Random Forest is an ensemble learner. It is a collection of many other learners. In this case many of the same kind of learner.