Ensemble Methods

Basics

1

Ensemble Methods

- Several models are trained to solve the task
- At application time, on a new instance, we combine their "votes" to make the final prediction.
- Important we have a diverse set of "voters"
- Main Questions:
  - How do we obtain several models from same data
  - How do we combine the votes

2

Why use Ensembles?

- Why do we vote?
- Two ways of presenting the intuition
  - A Theorem (though based on an impractical scenario)
  - Adding to complexity of models

3

Scenario

- Binary classification task
- Lots of learners (base models) – n
- Weak learners sufficient
  - Better than tossing coins (random guess)
    - accuracy > 0.5
    - i.e., error rate < 0.5
  - For each model, error rate = e
- Each model is independent (i.e., errors are not correlated).
  - What if they were all identical?

4

Probability of k errors

- On a random instance, assuming independence, probability of error for sequence
  - 1st: ✓ 2nd: x 3rd: x …
  - k errors and (n-k) correct.
- How many such sequences with k errors

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad = nCk$$

5

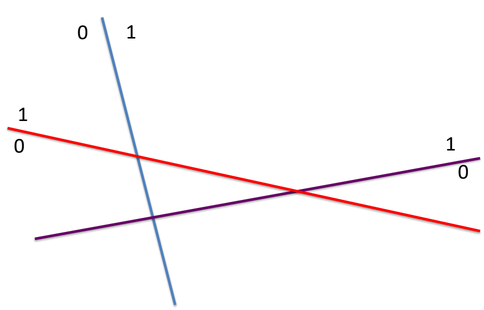
Errors rates

- $E_k = nCk\ e^k (1-e)^{n-k}$
- Given n, majority?
  - Examples n=12 → 6 & n=13 → 7
- Floor((n+1)/2)
- $E_{ensemble} = \sum_{K=Floor((n+1)/2)}^{n} nCk\ e^k (1-e)^{n-k}$

n = 25; e = 0.35 → error rate = 0.06

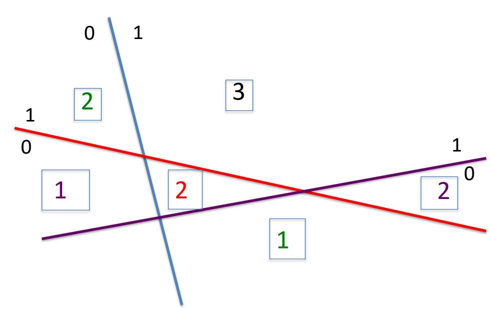
6

Combination 3 Linear Classifiers



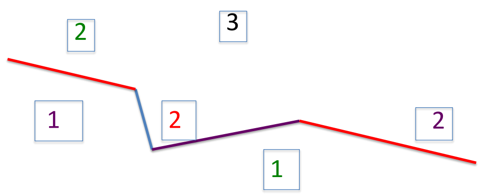
7

Majority Voting



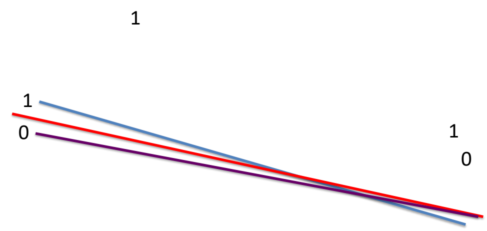
8

Separating Boundary



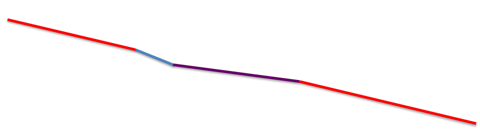
9

Majority Voting



10

Separating Boundary



11

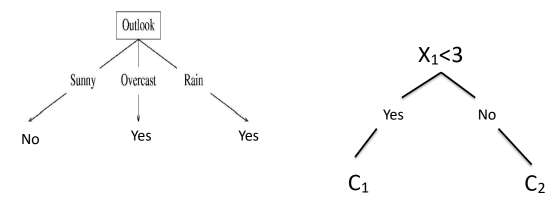
How do We Obtain Different Models

- Train Different models – SVM, MLP, …
  - Usually not much gain
  - Different Initializations, network topology
- Creating Different Training Sets
  - Bagging and Boosting
- Manipulating Input Features/Attributes
  - Random Forests
- Ensemble methods and Unstable classifiers
  - ML methods with high variance

12

Decision Stumps

- Depth = 1. Only one question asked. Root + leaves.



13