Ensemble Exercises - Part 1

Summary

The exercises here are mostly conceptual. You will see some practical exercises next time.

Exercise 1

Explain in your own words: What is an ML ensemble method? Why does it work?

Exercise 2

In ensemble methods, we want base-learners to be diverse. Why do we want divesrity? How can we approach it?

Exercise 3

Assume that a classification ensemble is composed of 5 base-learners, each of which is iid and correct with probability $p > 0.5$.

How do you calculate the probability that a majority vote from this ensemble gives the correct answer?

Exercise 4

We have a classification problem with two classes Red and Green. Suppose we have traind 10 classifiers that produce estimates $p(\mathrm{Red}|X)$: 0.1, 0.15, 0.2, 0.2, 0.55, 0.6, 0.6, 0.65, 0.7, 0.75

Use the two common voting approaches (hard- and soft-voting) and compute what is the final classification under each of these two approaches.

Exercise 5

Voting classifiers calculate their final output from the combined outputs of the individual base learners. The combined output can be calculated in different ways (with hard- and soft-voting being the most common).

Assume an ensemble H of m base classifiers, h_m (with $m = 1 \ldots M$). Given some input x, each classifier h_m predicts class probabilities for each of K output classes, so that $h_{mk}(x)$, and $\sum_{k=1}^{K} h_{mk}(x) = 1$.

The list below shows some different functions that can be used to combine the base-learner outputs to calculate the final output $\hat{y}$ of the whole voting classifier for input x.

What does each one do? What is the effect of using each?

1. $\hat{y} = \arg\max_k 1/M \sum_{m=1}^{M} h_{mk}(x)$
2. $\hat{y} = \arg\max_k \sum_{m=1}^{M} w_m h_{mk}(x)$, (here w is a weight, with $w_m \geq 0, \sum_{m=1}^{M} w_m = 1$)
3. $\hat{y} = \arg\max_k \mathrm{median}_m h_{mk}(x)$, ($\arg\max_k$ of median_m of $h_{mk}(x)$)
4. $\hat{y} = \arg\max_k \mathrm{min_m} h_{mk}(x)$
5. $\hat{y} = \arg\max_k \mathrm{max_m} h_{mk}(x)$
6. $\hat{y} = \arg\max_k \prod_m h_{mk}(x)$

Exercise 6

We want to derive the probability that a given observation is part of a bootstrap sample. Suppose that we obtain a bootstrap sample from a set of n observations.

- What is the probability that the first bootstrap observation is not the jth observation from the original sample? Justify your answer.
- What is the probability that the second bootstrap observation is not the jth observation from the original sample?
- Argue that the probability that the jth observation is not in the bootstrap sample is $(1 - 1/n)^n$
- When n = 5, what is the probability that the jth observation is in the bootstrap sample?
- When n = 100, what is the probability that the jth observation is in the bootstrap sample?
- When n = 1000, what is the probability that the jth observation is in the bootstrap sample?
- Create a plot that displays, for each integer value of n from 1 to 10000, the probability that the jth observation is in the bootstrap sample. Comment on what you observe.
- Now investigate numerically the probability that a bootstrap sample of size n = 100 contains the jth observation. Here j = 4. Repeatedly create bootstrap samples, and each time record whether or not the fourth observation is contained in the bootstrap sample. Comment on the obtained results.