DS102 Discussion 6

October 13, 2021

1. **Bootstrap and the Sample Maximum**

   Let $X_1, X_2, ..., X_n$ represent i.i.d draws from a Uniform$[0,1]$ distribution. We wish to use the bootstrap to understand the sampling distribution of the maximum,

   $$M_n = \max\{X_1, X_2, ..., X_n\}$$

   We will use $X_1^*, X_2^*, ..., X_n^*$ to denote the bootstrap resamples.

   (a) *Finding the distribution of the Sample Maximum*
   Compute $\mathbb{P}\left[M_n \leq t\right]$. Use this to compute the density of M_n.

(b) *Accuracy of Bootstrap Max estimates*

Let $M_n^* = \max\{X_1^*, X_2^*, ..., X_n^*\}$. Find $\mathbb{P}\left[M_n^* = M_n\right]$.

(c) *Quality of Bootstrap approximation of M_n*

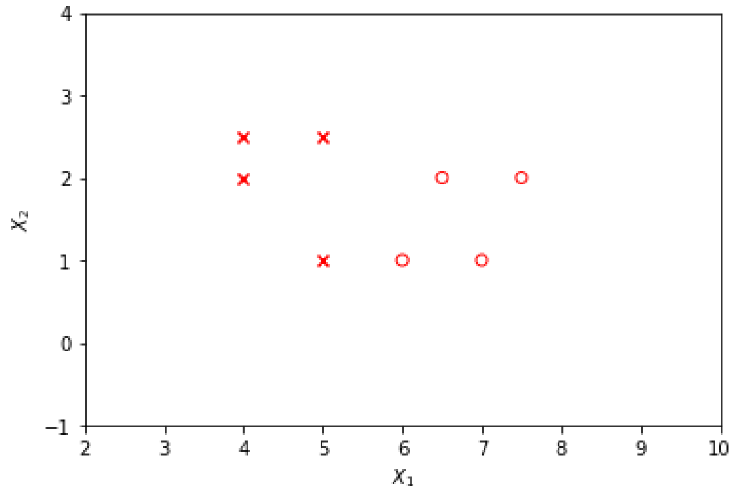
Is the distribution of M_n^* a good approximation for the distribution of M_n? Why is this result to be expected?

Hint: Use the fact that $\lim_{n\to\infty}\left(1 - \frac{1}{n}\right)^n = e^{-1}$.

2. **Bootstrap for Classification Models**

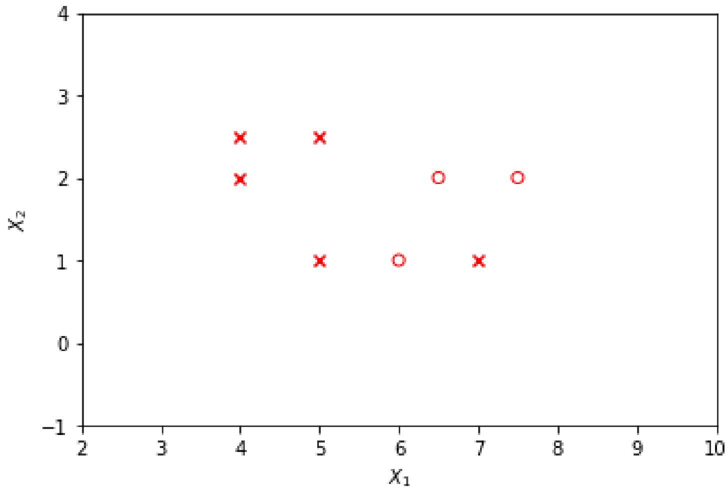
   (a) *Identifying Decision Boundaries*

   We have the following data set of two classes: X and O. Draw two plausible decision boundaries corresponding to a Logistic Regression classifier and a Decision Tree classifier overlaid on the scatter plot. Assume the Decision Tree is trained without a depth limit.



(b) *Changing Decision Boundaries with Data Shifts*

Now, we flip the data point at $(7, 1)$ from O to X, as shown in the following scatter plot. Once again, draw two plausible decision boundaries corresponding to a Logistic Regression classifier and a Decision Tree classifier overlaid on the scatter plot. How do the decision boundaries of each classifier change?



(c) *Data Shifts and Bootstrapping*

Based on your answers to (a) and (b), would bootstrapping provide a viable way to estimate the uncertainty of Logistic Regression and Decision Tree models? For which model would bootstrapping perform worse?

3. **Trees, Forests, Bias, and Variance**

Recall that we can express the Frequentist Risk of a decision procedure $\delta(x)$ with respect to parameter θ as:

$$R(\theta) = \mathbb{E}\left[(\delta(x) - \theta)^2\right] = \underbrace{\mathbb{E}\left[(\delta(x) - \mathbb{E}[\delta(x)])^2\right]}_{\text{Variance of } \delta(x)} + \underbrace{(\mathbb{E}[\delta(x)] - \theta)^2}_{\text{Bias}^2 \text{ of } \delta(x)}$$

If our decision is a prediction for y that we call $\hat{y}$ and $\delta(x) = \hat{y}(x)$, then we can re-write the above expression as:

$$E[(\hat{y}(x) - y)^2] = \underbrace{E\left[(\hat{y}(x) - E[\hat{y}(x)])^2\right]}_{\text{Variance of prediction } \hat{y}(x)} + \underbrace{(E[\hat{y}(x)] - y)^2}_{\text{Bias}^2 \text{ of } \hat{y}(x)}$$

In this question, we will consider the bias-variance decomposition for two non-parametric models: Decision Trees and Random Forests.

(a) *Bias-Variance Decomposition for Decision Trees*

Consider a Decision Tree trained without a limit on depth. Describe this model's bias and variance.

(b) *Bias-Variance Decomposition for Random Forests*

Compare a Random Forest's bias and variance to those of a Decision Tree. Which model would you expect to generalize better to unseen data and why?