UNIVERSITY OF EDINBURGH
Business School

Predictive Analytics and Modelling of Data

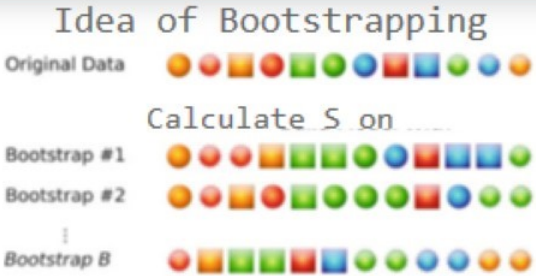
CMSE11428 (2020-2021)

Dr Xuefei Lu
The University of Edinburgh Business School

Bootstrapping

Bootstrapping: Idea

- The bootstrap is a general tool for assessing statistical accuracy. The basic idea of Bootstrapping is to randomly draw datasets with replacement from the data, each sample the same size as the original sample.



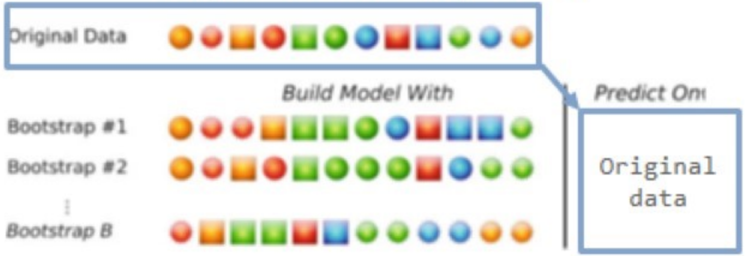
- That is, selecting n samples uniformly from a dataset of size n. Since we are doing replacement and we are sampling as many data points as are in the initial dataset, we can have repetition of some observations, while others will not appear.
- The above sampling step is done B times ($B = 100$ say), producing B bootstrap datasets.
- Then we calculate the statistical quantity $S(\boldsymbol{x}^b)$ (e.g. mean) based on each bootstrap sample $\boldsymbol{x}^b$, $b = 1 \ldots B$.
- $S(\boldsymbol{x}^b)$, $b = 1 \ldots B$ are used to assess the statistical accuracy of S. (e.g. the uncertainty of estimated mean based on a given finite sample)

bootstrapping

- Selecting n samples uniformly from a dataset of size n.
- Repeat B times to produce B bootstrap datasets. **Then we refit the model to each of the bootstrap datasets, and examine the behavior of the fits over the B replications.**

$$\widehat{err}_{boot} = \frac{1}{B}\frac{1}{n}\sum_{b=1}^{B}\sum_{i=1}^{n} error(y_i, \hat{f}^b(x_i))$$

Simple Bootstrap error

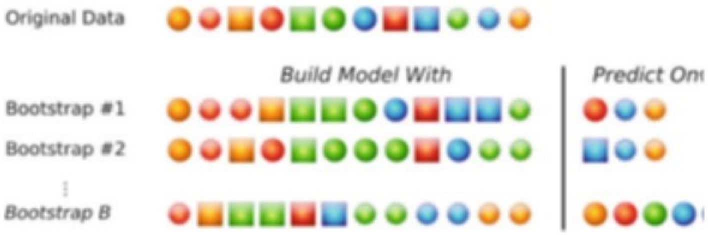


However, the above results are quite biased, as we used the same observations that we used in the bootstrap to construct our model. Therefore, our result is too optimistic about the model we have trained.

bootstrapping

- Selecting n samples uniformly from a dataset of size n. The observations do not appearing in this sample are called the **out-of-bag samples**.
- Repeat B times to produce B bootstrap datasets. **Then we refit the model to each of the bootstrap datasets, and examine the behavior of the fits over the B replications.**

leave-one out Bootstrap error



$$\widehat{err}_{LOOBS} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{|C^{-1}|}\sum_{b\in C^{-1}} error(y_i, \hat{f}^{b}(x_i))$$

Here C^{-1} is the set of indices of the bootstrap samples b that do not contain observation i, and $|C^{-1}|$ the number of such samples.

The leave-one out bootstrap solves the overfitting problem suffered by $\widehat{err}_{boot}$, but has the training-set-size bias like cross-validation.

bootstrapping

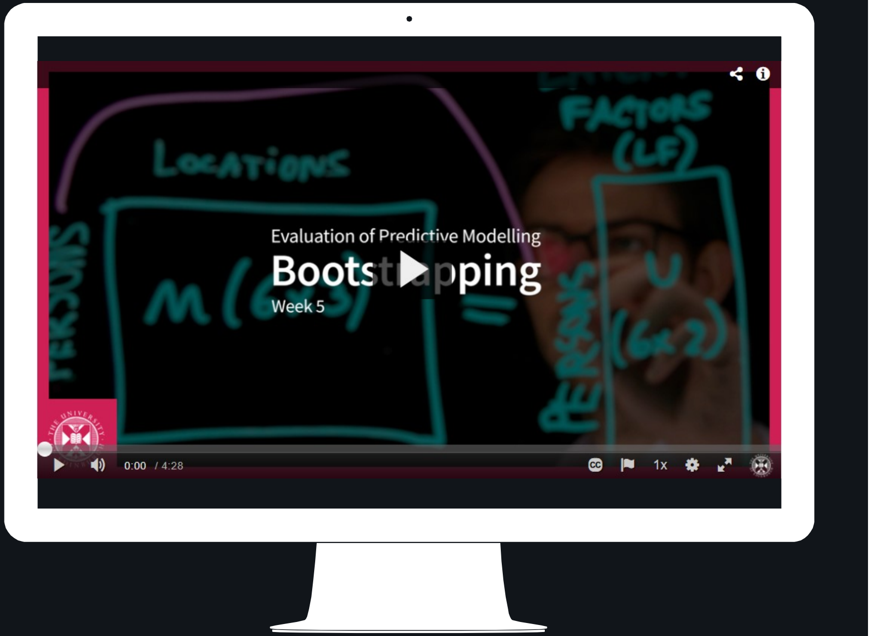
- The .632 method or .632 bootstrapping:

$$\widehat{err}_{.632} = \mathbf{0.368}\, \widehat{err}_{\boldsymbol{boot}} + \ \mathbf{0.632}\, \widehat{err}_{\boldsymbol{LOOBS}}$$

- Intuitively it pulls the leave-one out bootstrap estimate $\widehat{err}_{\boldsymbol{LOOBS}}$ down toward the training error rate, and hence reduces its upward bias.

- This reweighing of the metric can be taken into account when calculating, for example, the err can be accuracy or AUC of a predictive model for every bootstrap we run.

Bootstrapping



- Watch the video via the following link
- https://media.ed.ac.uk/media/Bootstrapping/0_6lrfpdpz/114521421

- Please study and try the following exercise
- 16 - How_to_code_for_bootstrapping.ipynb
- 17 - Activity_10_coding_bootstrapping-solution.ipynb

Activity: Coding bootstrapping

“

- # Discussion:
- Now that you have practised applying cross-validation, reflect on your work. You should review the results from your cross-validation, and check what the results were from just applying a 70/30 split.
- Consider and discuss the following:
  - □ What have you learned from applying cross-validation?
  - □ What is the difference between a training and test setup, shuffling, etc.?
  - □ What impact does shuffling have on the evaluation metrics?
  - □ Are your results robust and generalisable?

UNIVERSITY OF EDINBURGH
Business School