Random Forest

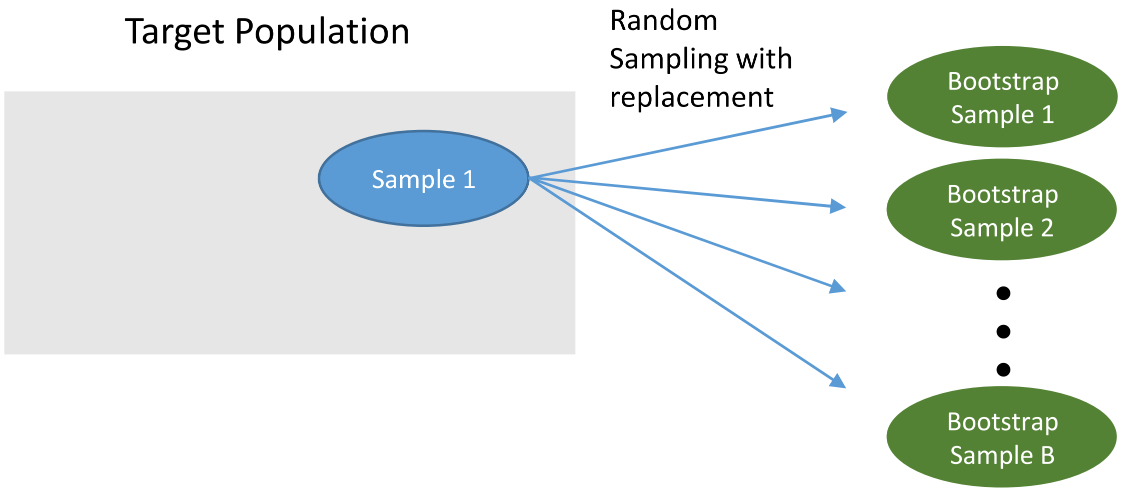
An Ensemble Machine Learning Technique

Part 1: Bootstrap Design vs Bagging Design

References

- Gareth James et. al. (2021) An Introduction to Statistical Learning. 2nd ed.
  - Sections in Chapter 8 explains Bagging and Random Forest.
  - eTextbook download: https://www.statlearning.com
- Chew C.H. (2022) Artificial Intelligence, Analytics and Data Science, Vol. 2.
  - Est. Q4 2022.
- Research Papers
  - Breiman (1996a) Bagging Predictors. *Machine Learning, 24, 123-140.*
  - Breiman (1996b) Heuristics of Instability and Stabilization in Model Selection. *The Annals of Statistics, Vol. 24, No. 6.*
  - Breiman (2001) Random Forests. *Machine Learning, 45, 5–32.*

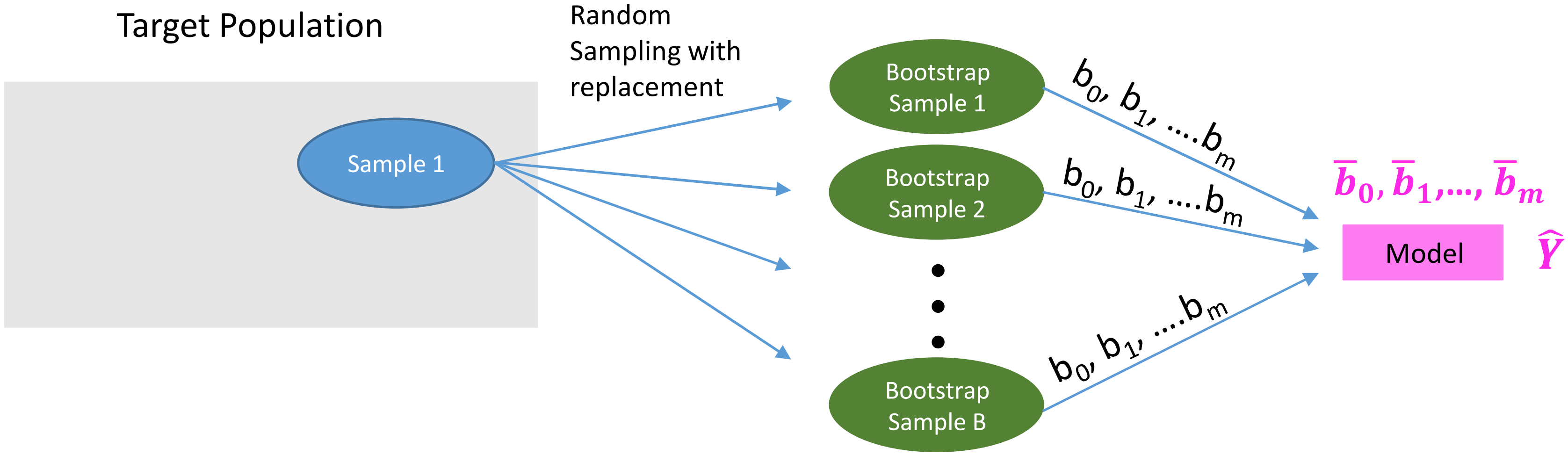
Bootstrap Design for Linear Regression Model



- Generate B bootstrap samples.

Sources: Chew C.H. (2022) AI, Analytics and Data Science, Vol. 2, Chap. 8.

Bootstrap Design for Linear Regression Model



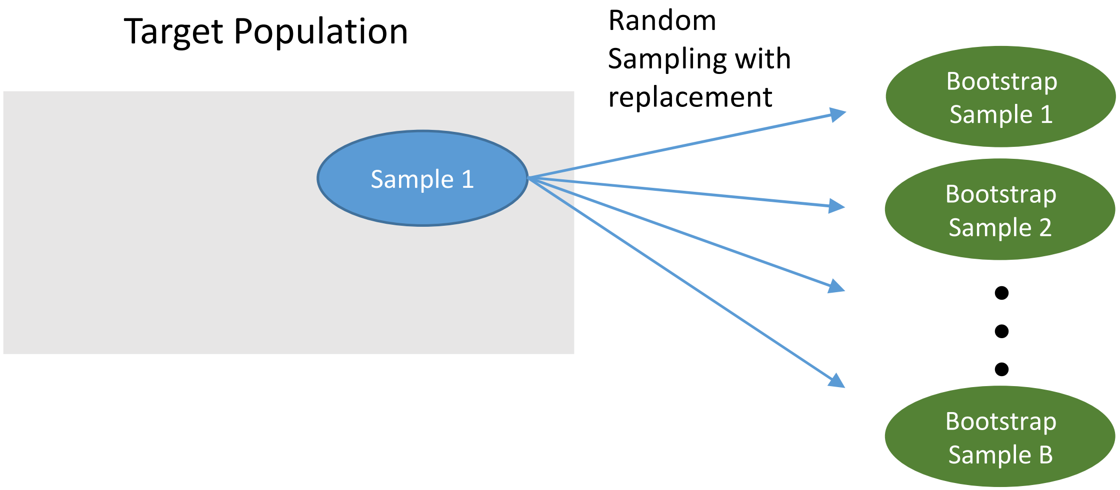
- Each bootstrap sample resulted in a different set of model coefficients b_0, b_1, ….b_m
- Take the mean of each model coefficients as the bootstrap estimate of the model coefficients for only one linear regression model.
  - For i = 0, 1,…m:

$$\overline{b}_i = \frac{\sum_{s=1}^{B} b_{i,s}}{B}$$

- Use that one linear regression model for prediction. i.e. get only one prediction.
- Same process applies for any other type of model.

Sources: Chew C.H. (2022) AI, Analytics and Data Science, Vol. 2, Chap. 8.

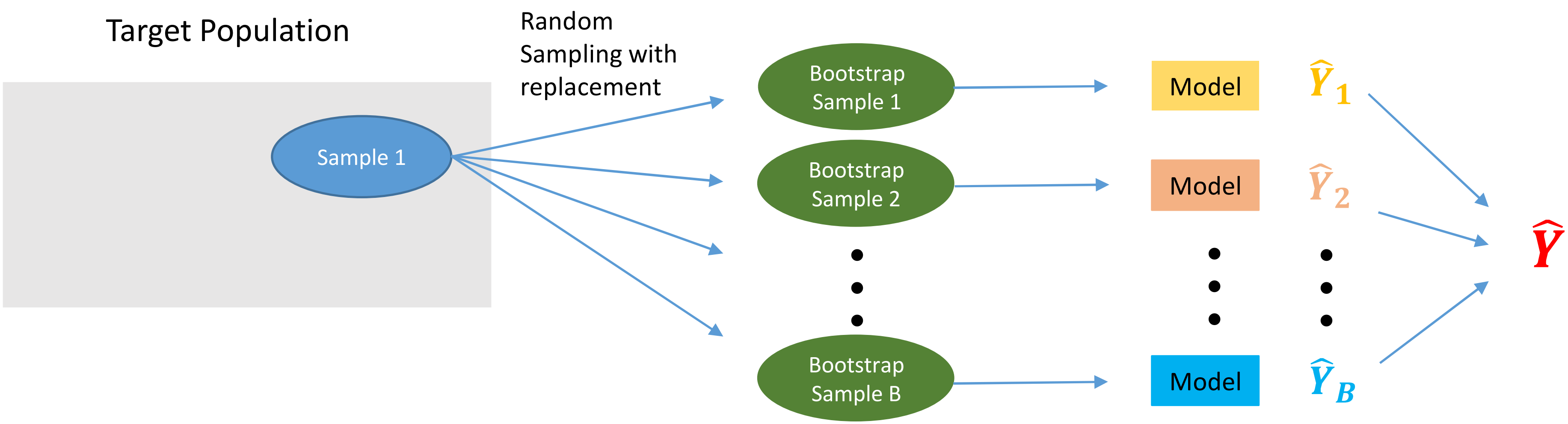
Bootstrap Aggregating (Bagging)



- Generate B bootstrap samples.

Sources: Chew C.H. (2022) AI, Analytics and Data Science, Vol. 2, Chap. 8.

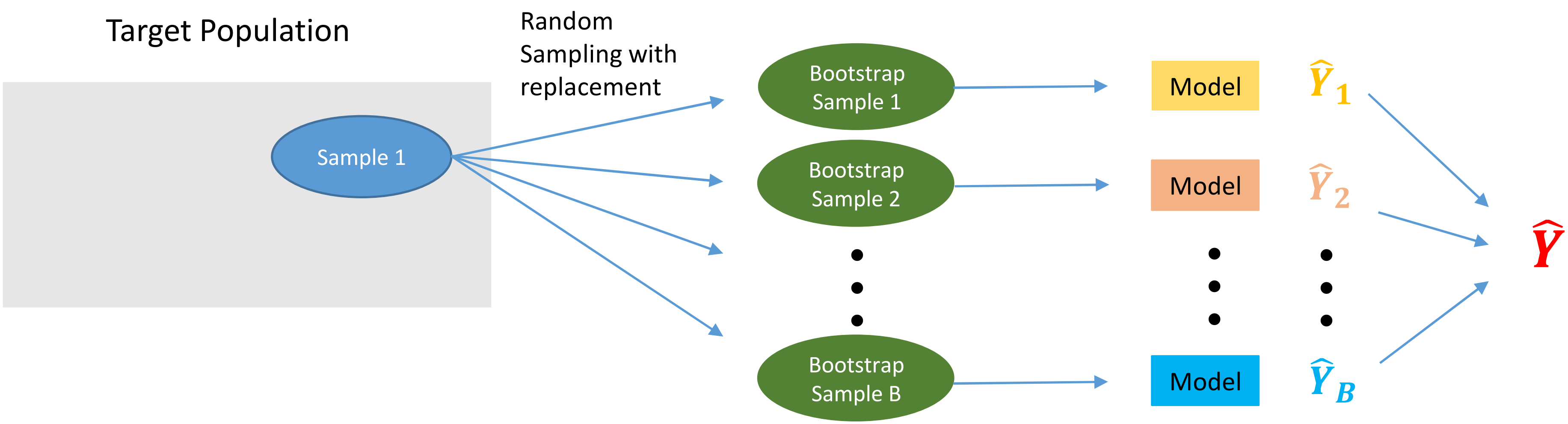
Bootstrap Aggregating (Bagging) Design



- Each bootstrap sample is used to create one model, unlike Bootstrap Design.
- The number of Bootstrap samples = number of models.
- We can choose to use any model (e.g. Linear Reg, Logistic Reg) but most common choice is CART.
- We get B predictions from B models, unlike Bootstrap Design.
- Finally aggregate the results to produce the final prediction.
- Experiments done by Breiman [results reported in 1996 paper]:
  - What should the Bootstrap sample size be? Same size as original sample (n)? Beneficial to use bigger sample?
  - How many Bootstrap samples should we generate (B)? Beneficial to generate a lot?

Sources: Chew C.H. (2022) AI, Analytics and Data Science, Vol. 2, Chap. 8 & Breiman (1996a)

Bootstrap Aggregating (Bagging) Design



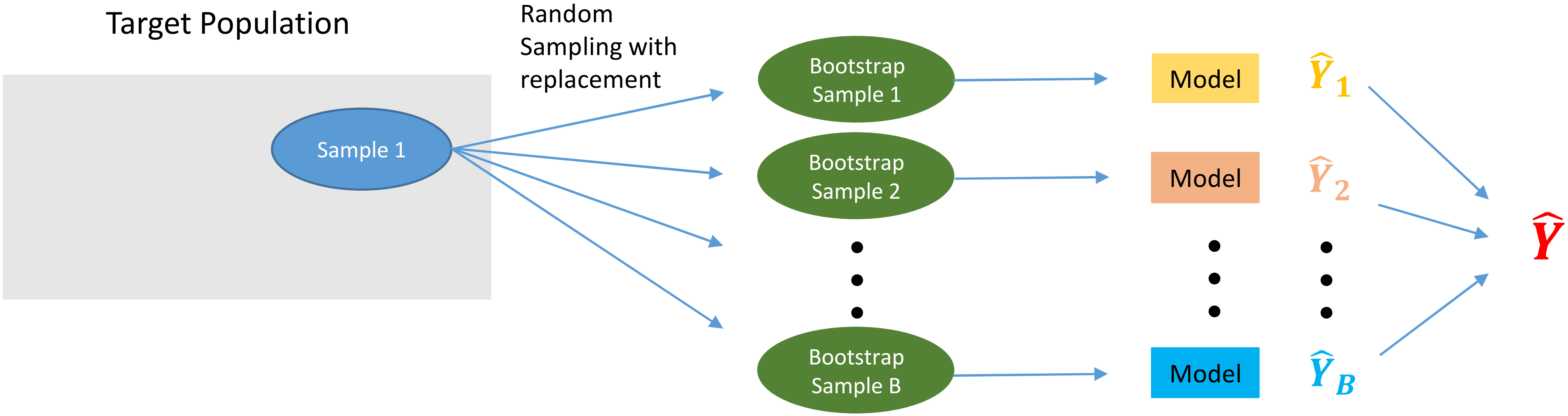
- Each bootstrap sample has the same size (n) as the original Sample 1.
  - Doubling bootstrap sample size to 2n

    - "There is no improvement in accuracy." --Breiman (1996a).
- B is chosen to be a large number. Breiman (1996a): 50 for Classification Tree, 25 for Regression Tree.
  - *"This does not mean that 50 or 25 were necessary or sufficient, but simply that they seemed reasonable. My sense of it is that fewer are required when y is numerical and more are required with an increasing number of classes." -- Breiman (1996a).*
  - *"…in practice we use a value of B sufficiently large for the error rate to have settled down."*

    -- ISLR

Sources: Chew C.H. (2022) AI, Analytics and Data Science, Vol. 2, Chap. 8 & Breiman (1996a)

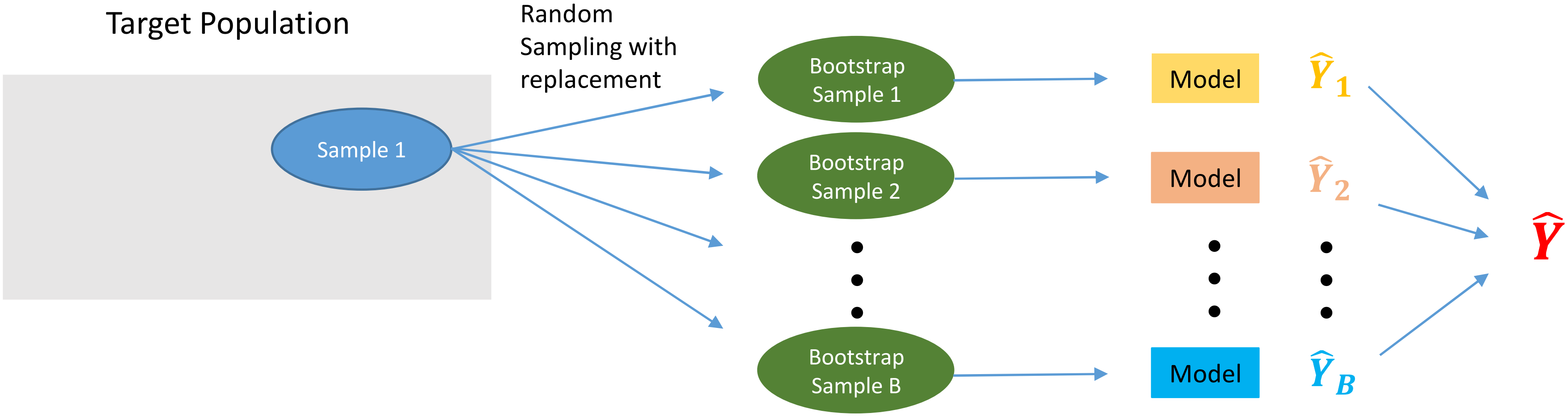
Bootstrap Aggregating (Bagging) Predicted Y



- Bagging Prediction of Y:
  - For continuous Y:
    - Take the average of $\hat{Y}_1, \hat{Y}_2, \dots \hat{Y}_B$
  - For categorical Y:
    - Take the mode (i.e. majority vote) of $\hat{Y}_1, \hat{Y}_2, \dots \hat{Y}_B$

Sources: Chew C.H. (2022) AI, Analytics and Data Science, Vol. 2: Chap. 8 & Breiman (1996a)

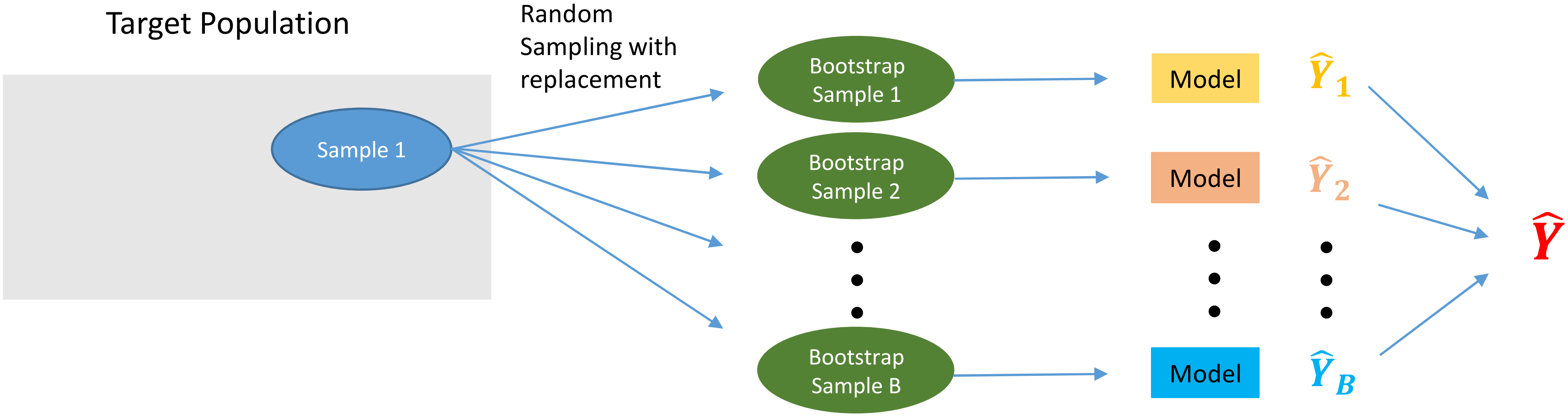
Bootstrap Aggregating (Bagging) Errors



- To get trainset error in Bagging, just take the average of the errors within the B Bootstrap samples.
- Q: How to get testset error in Bagging?

Sources: Chew C.H. (2020) AI, Analytics and Data Science, Vol. 2: Advanced Predictive Techniques, Chap. 8 & Breiman (1996a)

Bootstrap Aggregating (Bagging) Errors



- To get trainset error in Bagging, just take the average of the errors within the B Bootstrap samples.
- Q: How to get testset error in Bagging?
  - Breiman (1996a) did 90-10 train-test split from original sample 1 repeatedly (100 times).
  - A more efficient and natural testset procedure was subsequently devised in Breiman (2001) – OOB Error.

Sources: Chew C.H. (2020) AI, Analytics and Data Science, Vol. 2: Advanced Predictive Techniques, Chap. 8 & Breiman (1996a)

Example: Generate 3 bootstrap samples from a original data sample with 10 case IDs.

	A	B	C	D	E	F	G	H	I	J	K
1	Case ID	Range of Selection		Rand1	Bootstrap Sample 1		Rand2	Bootstrap Sample 2		Rand3	Bootstrap Sample 3
2	1	(0, 0.1]		0.6670843	7		0.04658482	1		0.7124438	8
3	2	(0.1, 0.2]		0.1060074	2		0.27705979	3		0.6669323	7
4	3	(0.2, 0.3]		0.2355851	3		0.35832777	4		0.674917	7
5	4	(0.3, 0.4]		0.3899567	4		0.10711809	2		0.1852939	2
6	5	(0.4, 0.5]		0.0840879	1		0.99876529	10		0.1711216	2
7	6	(0.5, 0.6]		0.5393161	6		0.28596224	3		0.081874	1
8	7	(0.6, 0.7]		0.58984	6		0.77837503	8		0.9558981	10
9	8	(0.7, 0.8]		0.938273	10		0.84775176	9		0.5442069	6
10	9	(0.8 0.9]		0.6780694	7		0.33486749	4		0.2299222	3
11	10	(0.9, 1]		0.3217754	4		0.04894023	1		0.5944013	6
12				Cases OOB:	5, 8, 9		Cases OOB:	5, 6, 7		Cases OOB:	4, 5, 9

- Bootstrap will naturally generate Out-of-Bag (OOB) sample for each bootstrap sample.
- OOB sample serve as the testset for that in-the-bag bootstrap sample.

Out-of-Bag (OOB) Error as the natural Testset Error in Bagging

- For a bootstrap sample of size n, the probability that case i in the original sample is not in a specific bootstrap sample:
  - P(case i is OOB) = $(1 – 1/n)^n$
  - If n = 10, $(1 – 1/n)^n \approx 0.3487\ldots$

> "each bagged tree makes use of around two-thirds of the observations" -- ISLR.

- $(1 – 1/n)^n \rightarrow 1/e$ as $n \rightarrow \infty$.
  - For 3 proofs, see https://socratic.org/questions/how-do-you-find-the-limit-of-1-1-x-x-as-x-approaches-infinity *[Optional. Skip if you are not interested in the math.]*

- Since $1/e \approx 0.367879\ldots \approx 0.3333\ldots = 1/3$
- Thus, as a convenient approximation, P(case i is OOB) ≈ 1/3.
- Since bootstrap samples are independent and identically distributed, approximately 1/3 of cases in original sample do not appear in bootstrap sample s and hence, those OOB cases could serve as a testset while bootstrap sample s serve as trainset.
- Get the mean OOB testset error.

Next – Performance of Bagging

- The performance of Bagging
  - When it worked well
  - When it did not improve accuracy
    - Leads to Random Forest…