SRSWOR

$1, 2, \ldots 10$

$\underbrace{\qquad\qquad}$

4

$\downarrow$

$8, \underset{2^{nd}}{3}, \underset{3^{rd}}{7}, \underset{4^{th}}{2}$

$\dfrac{1}{10} \quad \dfrac{}{9} \quad \dfrac{}{8} \quad \dfrac{}{7}$

SRSWR

$1, 2, \ldots 10$

$\underbrace{\qquad\qquad}$

4

$\downarrow$

$8, \underset{2^{nd}}{3}, \underset{3^{rd}}{8}, \underset{4^{th}}{2}$

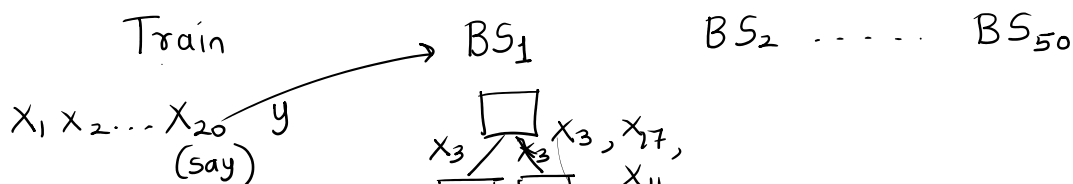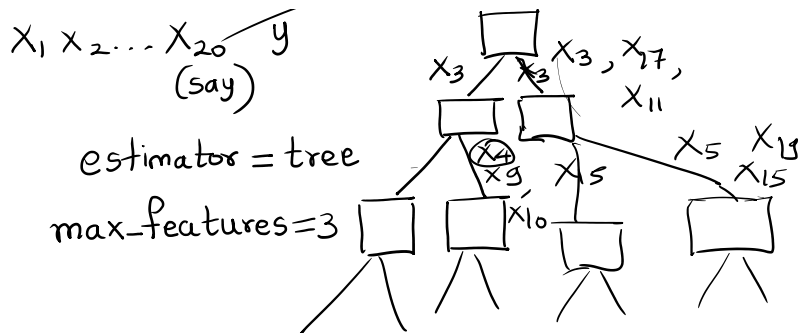$\dfrac{1}{10} \quad \dfrac{}{10} \quad \dfrac{}{10} \quad \dfrac{}{10}$

OOB Score (default setting)
— Classification ; accuracy score
— Regression ; R^2 score

Model	Params	Best Score Log Loss	Test Log Loss	Test Accuracy Score
SVM-Radial	{'SVM__gamma': 8.889, 'SVM__C': 6.667000000000001}	-0.19740990035900283	0.19761762450347778	0.929144385026738
Logistic	{'C': 0.001, 'l1_ratio': 0.0, 'penalty': 'l2'}	-0.16103620505179841	0.15490104275867858	0.9505347593582888
Tree	{'max_depth': 3, 'min_samples_leaf': 50, 'min_samples_split': 2}	-0.1678651753449944	0.15863970995495377	0.9505347593582888
Bagging with Logistic			0.9505347593582888	0.15497012508696711
Grid search of bagging with Logistic	{'n_estimators': 25}	-0.16106072762063564	0.9505347593582888	0.15500342982364854

Random Forest

Train $\longrightarrow BS_1 \qquad BS_2 \ldots BS_{50}$

$X_1 X_2 \ldots X_{20} \; y$
(say)

$X_3, X_3, X_3, X_{17}, X_{11}$

$X_1 \ X_2 \ldots X_{20} \ \ y$
(say)

estimator = tree

max_features = 3

X_3 X_3 X_3, X_{17}, X_{11}

X_4 X_9 X_5 X_{10} X_5 X_{14} X_{15}

max_features : *int, float or {"auto", "sqrt", "log2"}, default=None*
The number of features to consider when looking for the best split:

Sorted Importances Plot for Single Tree

Sorted Importances Plot for Random Forest

Sorted Importances Plot for Single Tree

Log loss =
-0.11424197097491764

Sorted Importances Plot for Random Forest

Log Loss =
-0.08728637173510187