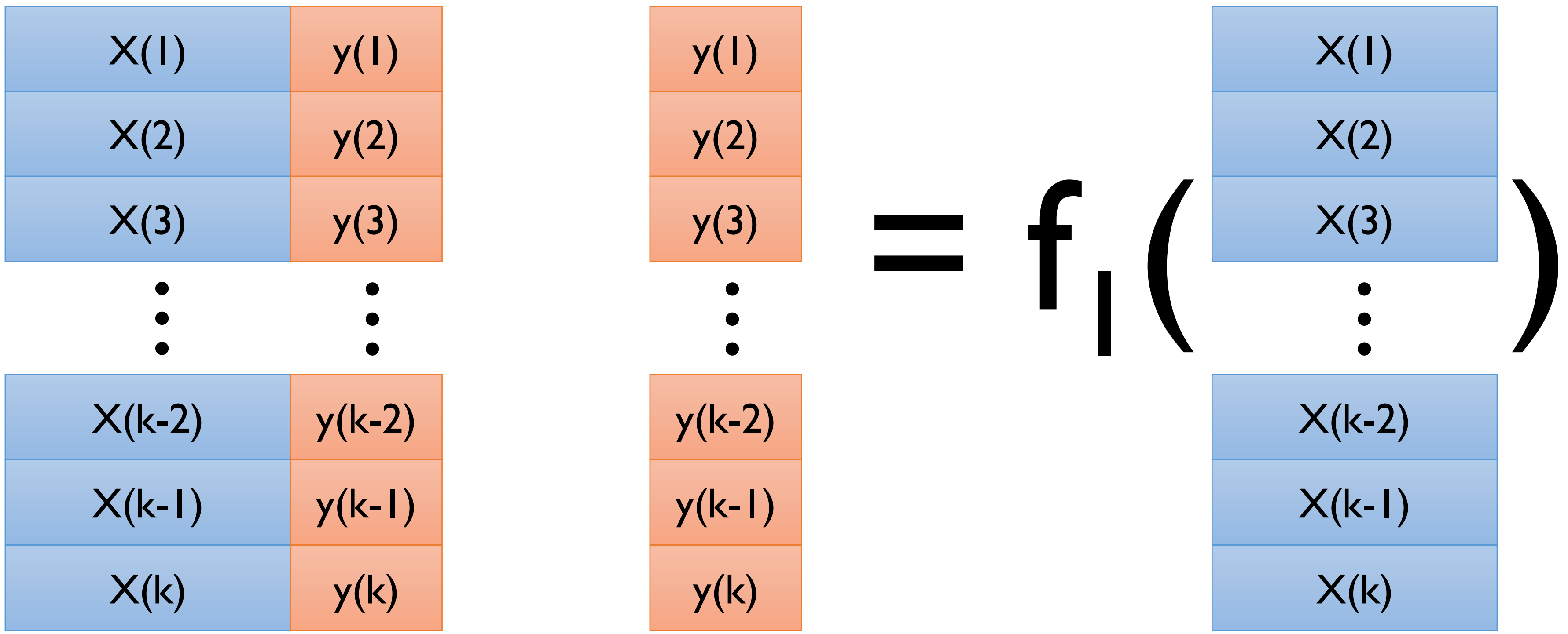
Ensemble Learning: Bagging

Pilsung Kang
School of Industrial Management Engineering
Korea University

Sampling without Replacement

- K-fold data split
  - ✓ Entire data is split into k blocks; each classifier is trained only on different subset of (k-1) blocks

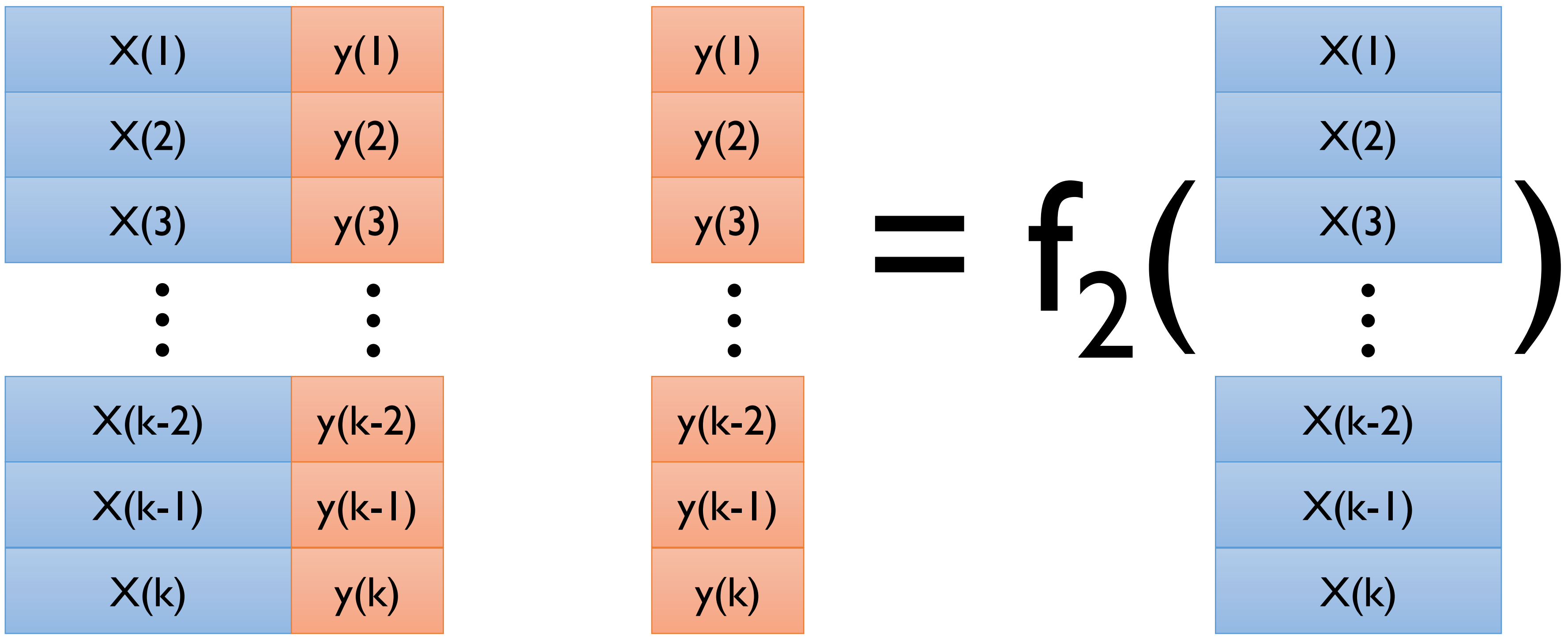


고려대학교 KOREA UNIVERSITY

DSBA Data Science & Business Analytics

Sampling without Replacement

- K-fold data split
  - ✓ Entire data is split into k blocks; each classifier is trained only on different subset of (k-1) blocks



Sampling without Replacement

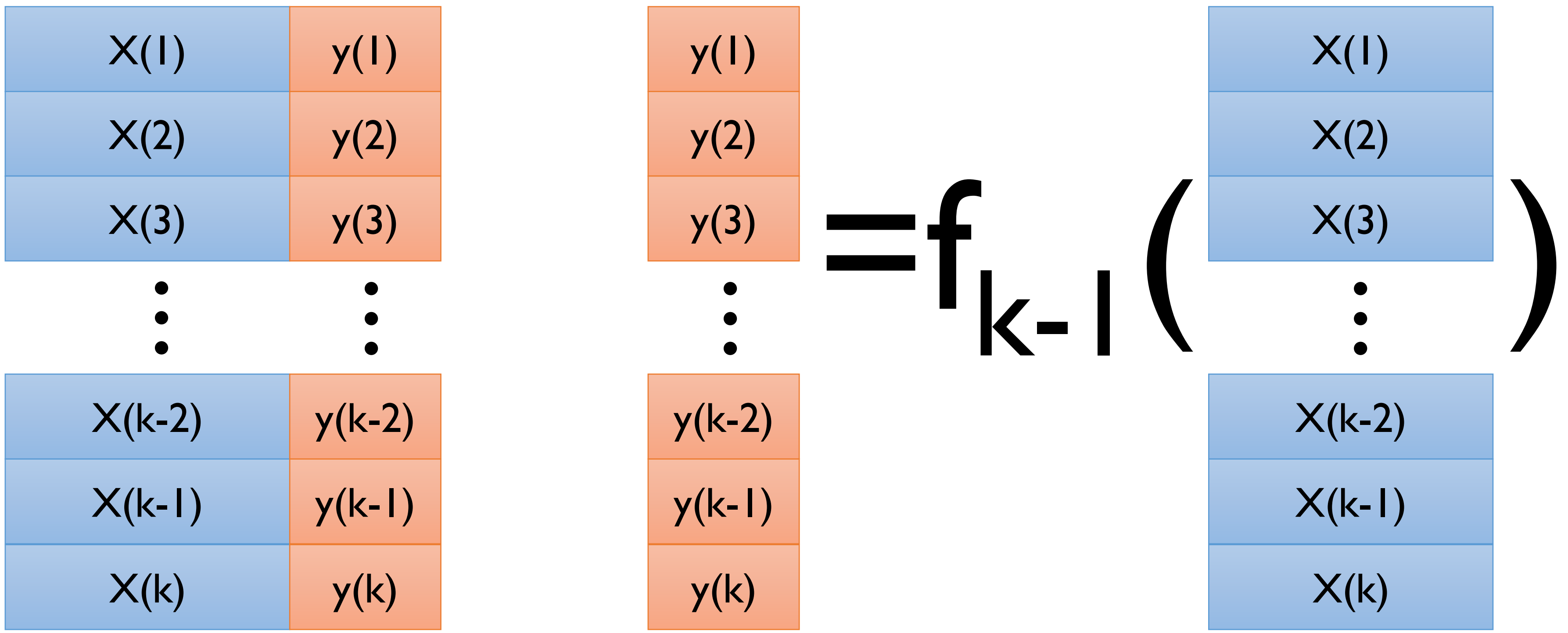
- K-fold data split
  - ✓ Entire data is split into k blocks; each classifier is trained only on different subset of (k-1) blocks

X(1)	y(1)
X(2)	y(2)
X(3)	y(3)
⋮	⋮
X(k-2)	y(k-2)
X(k-1)	y(k-1)
X(k)	y(k)

$$\begin{bmatrix} y(1) \\ y(2) \\ y(3) \\ \vdots \\ y(k-2) \\ y(k-1) \\ y(k) \end{bmatrix} = f_3\left(\begin{bmatrix} X(1) \\ X(2) \\ X(3) \\ \vdots \\ X(k-2) \\ X(k-1) \\ X(k) \end{bmatrix} \right)$$

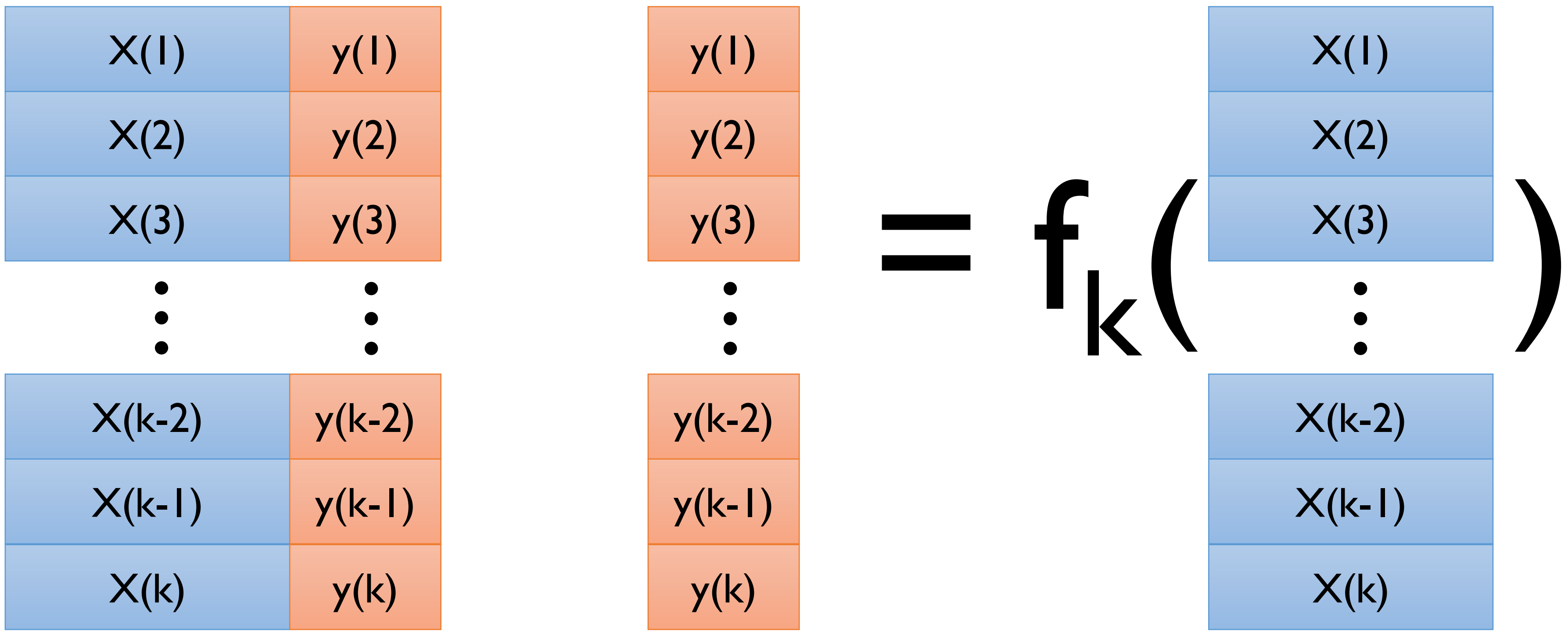
Sampling without Replacement

- K-fold data split
  - ✓ Entire data is split into k blocks; each classifier is trained only on different subset of (k-1) blocks



Sampling without Replacement

- K-fold data split
  - ✓ Entire data is split into k blocks; each classifier is trained only on different subset of (k-1) blocks



Sampling without Replacement

- K-fold data split
  - ✓ Entire data is split into k blocks; each classifier is trained only on different subset of (k-1) blocks
- Final output $\hat{y} = \delta\Big(f_1(\mathbf{x}), f_2(\mathbf{x}), \cdots, f_{k-1}(\mathbf{x}), f_k(\mathbf{x})\Big)$
  - ✓ $\delta(\cdot)$: An aggregation function of individual outputs (ex: simple average)

Bootstrap Aggregating: Bagging

Breiman (1996)

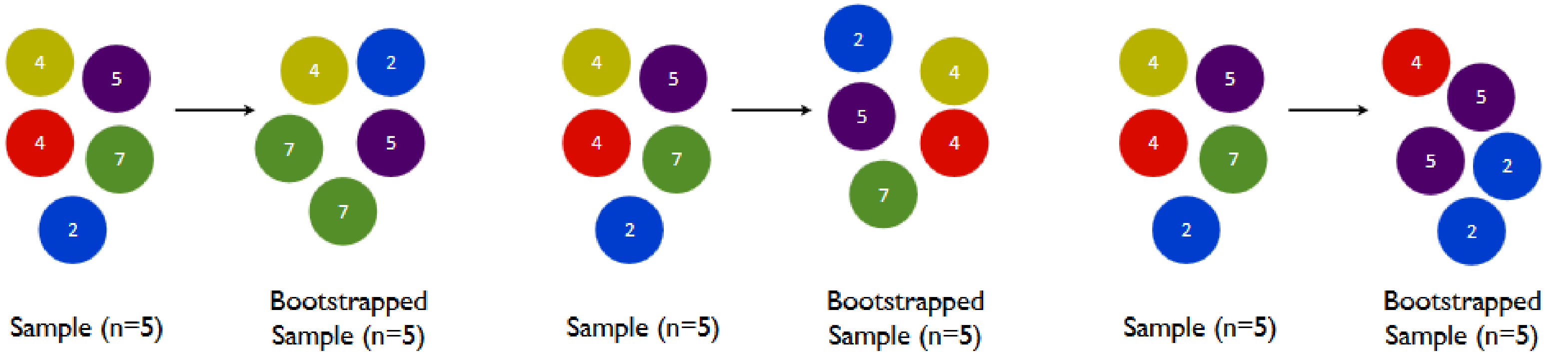
- Main Idea
  - ✓ Each member of the ensemble is constructed from a different training dataset
  - ✓ Each dataset is generated by sampling from the total *N* data examples, choosing *N* items uniformly at random with replacement
  - ✓ Each dataset sample is known as a bootstrap

Original Dataset		Bootstrap 1		Bootstrap 2		…	Bootstrap B	
x^1	y^1	x^3	y^3	x^7	y^7		x^9	y^9
x^2	y^2	x^6	y^6	x^1	y^1		x^5	y^5
x^3	y^3	x^2	y^2	x^{10}	y^{10}		x^2	y^2
x^4	y^4	x^{10}	y^{10}	x^1	y^1		x^4	y^4
x^5	y^5	x^8	y^8	x^8	y^8		x^7	y^7
x^6	y^6	x^7	y^7	x^6	y^6		x^2	y^2
x^7	y^7	x^7	y^7	x^2	y^2		x^5	y^5
x^8	y^8	x^3	y^3	x^6	y^6		x^{10}	y^{10}
x^9	y^9	x^2	y^2	x^4	y^4		x^8	y^8
x^{10}	y^{10}	x^7	y^7	x^9	y^9		x^2	y^2

Bootstrap Aggregating: Bagging

- Bagging: Bootstrapp Aggregating
  - ✓ Probability that an instance is not included in a bootstrap

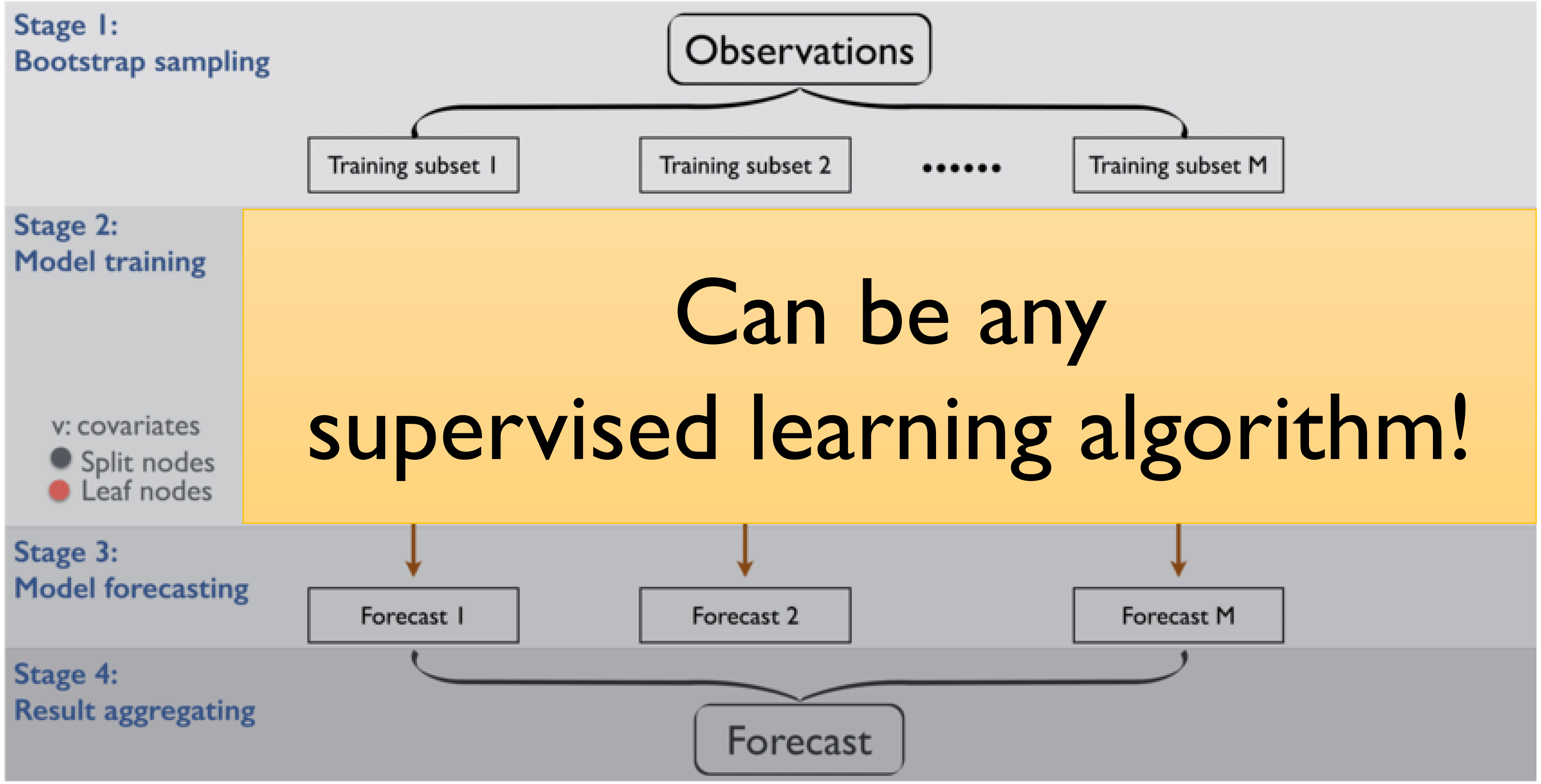
$$p = \left(1 - \frac{1}{N}\right)^N \rightarrow \lim_{N \to \infty} \left(1 - \frac{1}{N}\right)^N = e^{-1} = 0.368$$



  - ✓ Fits well with the models with low bias and high variance

Bootstrap Aggregating: Bagging

- Bagging with Decision Tree



Bootstrap Aggregating: Bagging

- Result Aggregating
  - ✓ For classification problem
    - Majority voting

$$\hat{y}_{Ensemble} = arg\max_i \left(\sum_{j=1}^{n} \delta(\hat{y}_j = i), \quad i \in \{0, 1\} \right)$$

Training Accuracy	Ensemble population	P(y=1) for a test instance	Predicted class label
0.80	Model 1	0.90	1
0.75	Model 2	0.92	1
0.88	Model 3	0.87	1
0.91	Model 4	0.34	0
0.77	Model 5	0.41	0
0.65	Model 6	0.84	1
0.95	Model 7	0.14	0
0.82	Model 8	0.32	0
0.78	Model 9	0.98	1
0.83	Model 10	0.57	1

$$\sum_{j=1}^{n} \delta(\hat{y}_j = 0) = 4$$

$$\sum_{j=1}^{n} \delta(\hat{y}_j = 1) = 6$$

$$\hat{y}_{Ensemble} = 1$$

Bootstrap Aggregating: Bagging

- Result Aggregating
  - ✓ For classification problem
    - Weighted voting (weight = training accuracy of individual models)

$$\hat{y}_{Ensemble} = arg\max_i \left(\frac{\sum_{j=1}^{n}(TrnAcc_j) \cdot \delta(\hat{y}_j = i)}{\sum_{j=1}^{n}(TrnAcc_j)}, \quad i \in \{0, 1\} \right)$$

Training Accuracy	Ensemble population	P(y=1) for a test instance	Predicted class label
0.80	Model 1	0.90	1
0.75	Model 2	0.92	1
0.88	Model 3	0.87	1
0.91	Model 4	0.34	0
0.77	Model 5	0.41	0
0.65	Model 6	0.84	1
0.95	Model 7	0.14	0
0.82	Model 8	0.32	0
0.78	Model 9	0.98	1
0.83	Model 10	0.57	1

$$\frac{\sum_{j=1}^{n}(TrnAcc_j) \cdot \delta(\hat{y}_j = 0)}{\sum_{j=1}^{n}(TrnAcc_j)} = 0.424$$

$$\frac{\sum_{j=1}^{n}(TrnAcc_j) \cdot \delta(\hat{y}_j = 1)}{\sum_{j=1}^{n}(TrnAcc_j)} = 0.576$$

$$\hat{y}_{Ensemble} = 1$$

Bootstrap Aggregating: Bagging

- Result Aggregating
  - ✓ For classification problem
    - Weighted voting (weight = predicted probability for each class)

$$\hat{y}_{Ensemble} = arg\max_i \left(\frac{1}{n}\sum_{j=1}^{n} P(y=i), \quad i \in \{0,1\}\right)$$

Training Accuracy	Ensemble population	P(y=1) for a test instance	Predicted class label
0.80	Model 1	0.90	1
0.75	Model 2	0.92	1
0.88	Model 3	0.87	1
0.91	Model 4	0.34	0
0.77	Model 5	0.41	0
0.65	Model 6	0.84	1
0.95	Model 7	0.14	0
0.82	Model 8	0.32	0
0.78	Model 9	0.98	1
0.83	Model 10	0.57	1

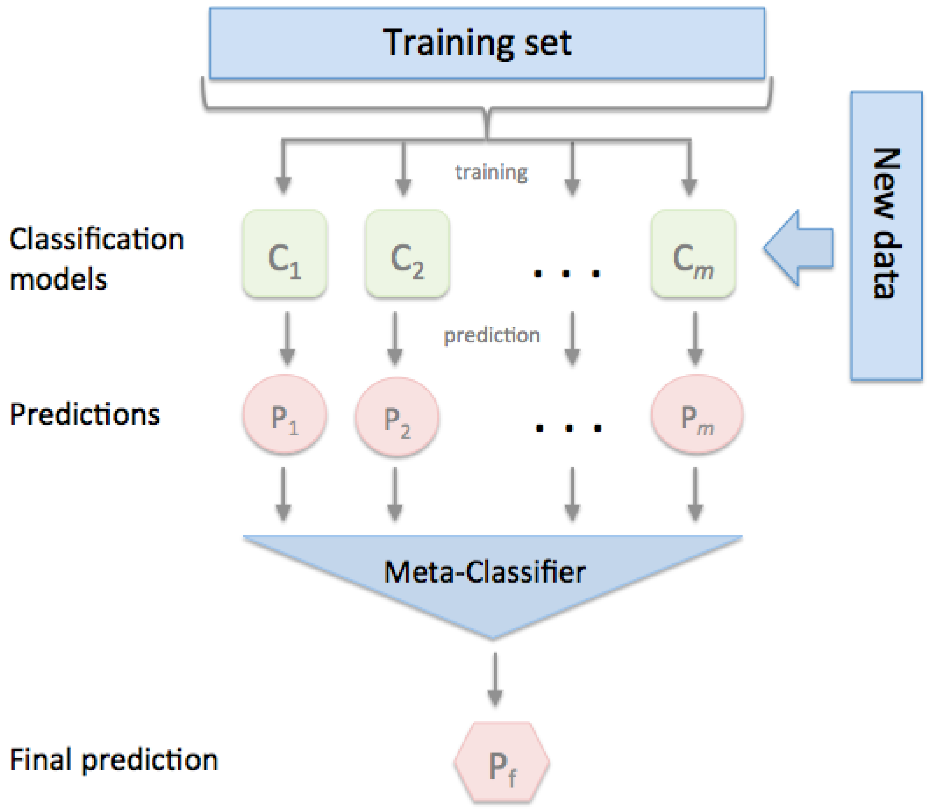
$$\frac{1}{n}\sum_{j=1}^{n} P(y=0) = 0.375$$

$$\frac{1}{n}\sum_{j=1}^{n} P(y=1) = 0.625$$

$$\hat{y}_{Ensemble} = 1$$

Bootstrap Aggregating: Bagging

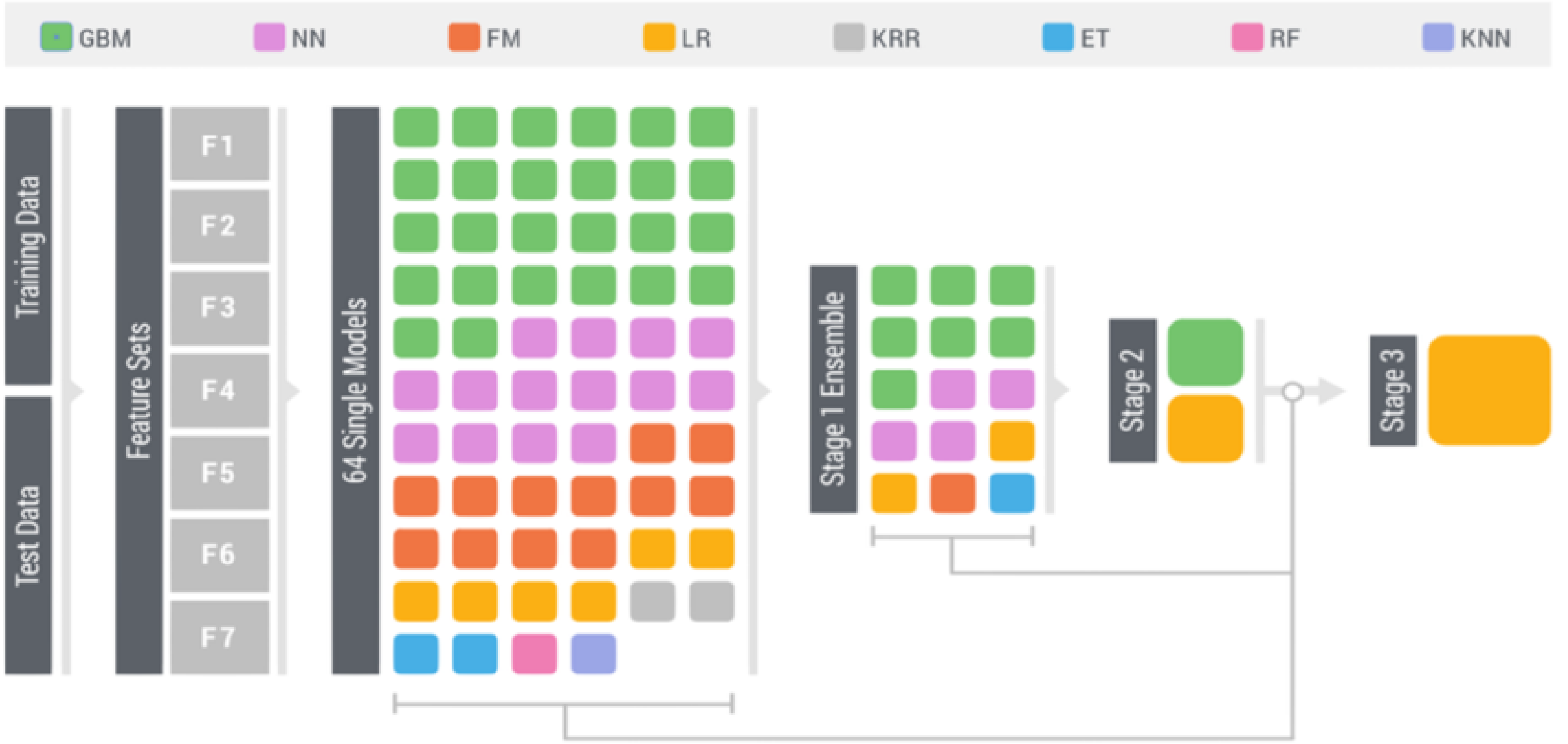
- Result Aggregating: Stacking
  - ✓ Use another prediction model to aggregate the results
    - Input: Predictions made by ensemble members
    - Target: Actual true label



https://rasbt.github.io/mlxtend/user_guide/classifier/StackingClassifier/

Bootstrap Aggregating: Bagging

- Result Aggregating: Stacking
  - ✓ The winner of KDD-cup 2015
    - MOOC dropout prediction



•Jeong-Yoon Lee, Winning Data Science Competitions

Bootstrap Aggregating: Bagging

- Bagging: Algorithm

Algorithm 1 Bagging

Input: Required ensemble size T
Input: Training set $S = \{(x_1, y_1), (x_2, y_2), ..., (x_N, y_N)\}$
for $t = 1$ to T **do**
  Build a dataset S_t, by sampling N items, randomly *with replacement* from S.
  Train a model h_t using S_t, and add it to the ensemble.
end for
For a new testing point (x', y'),
If model outputs are continuous, combine them by averaging.
If model outputs are class labels, combine them by voting.

Bootstrap Aggregating: Bagging

- Bagging: Illustration

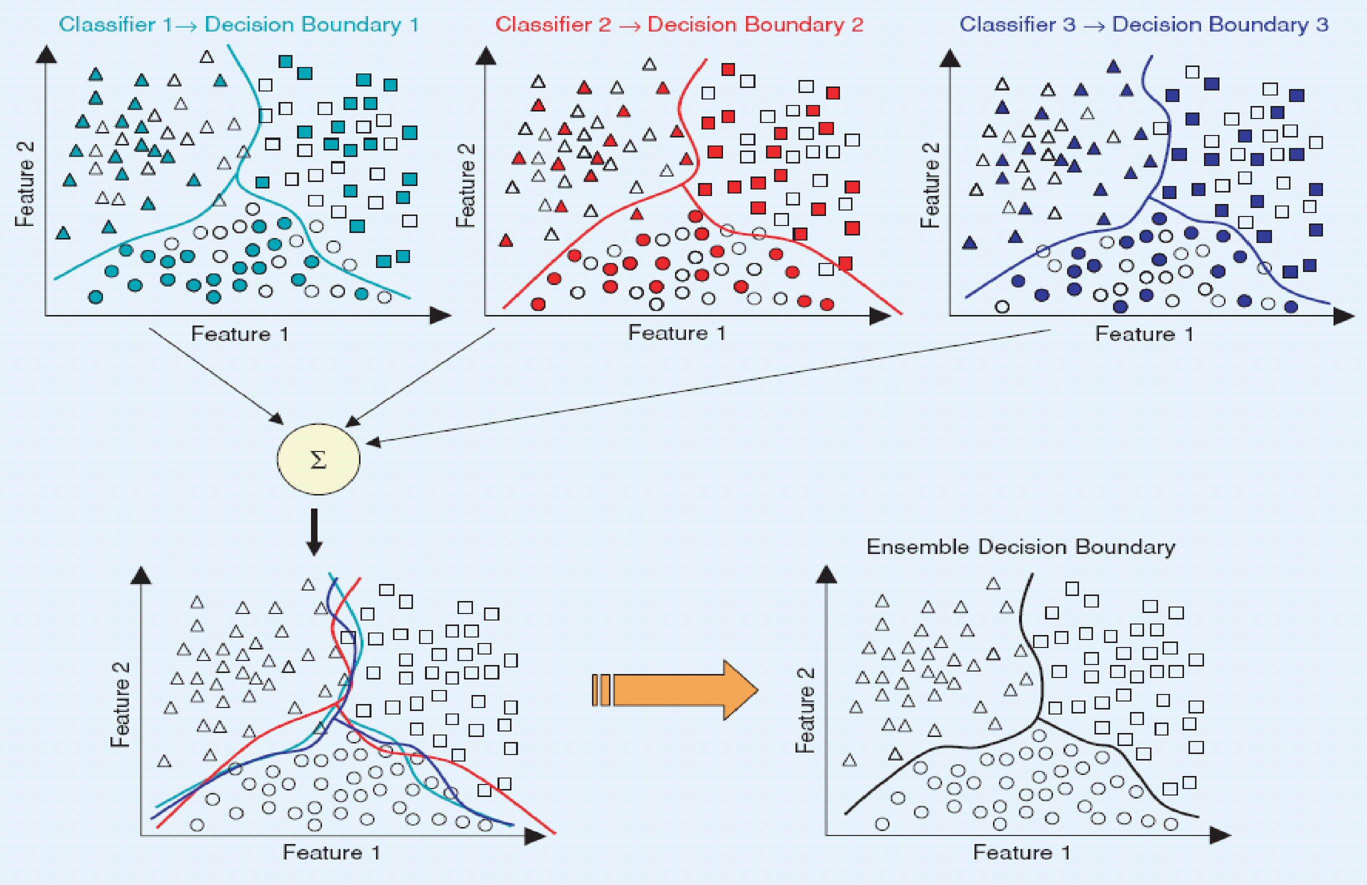


Figure 3. Combining classifiers that are trained on different subsets of the training data.

Bootstrap Aggregating: Bagging

- Aggregation examples

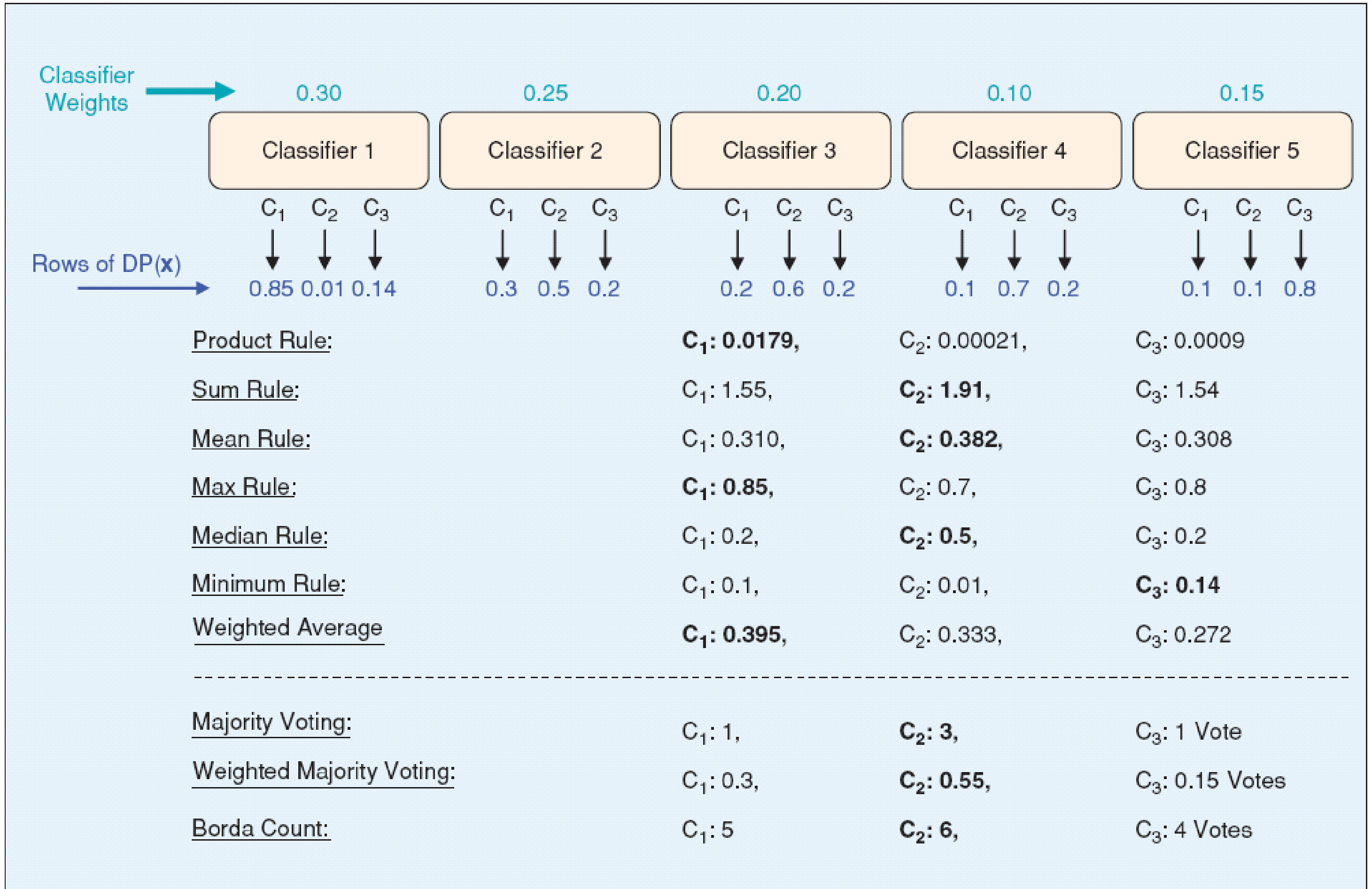
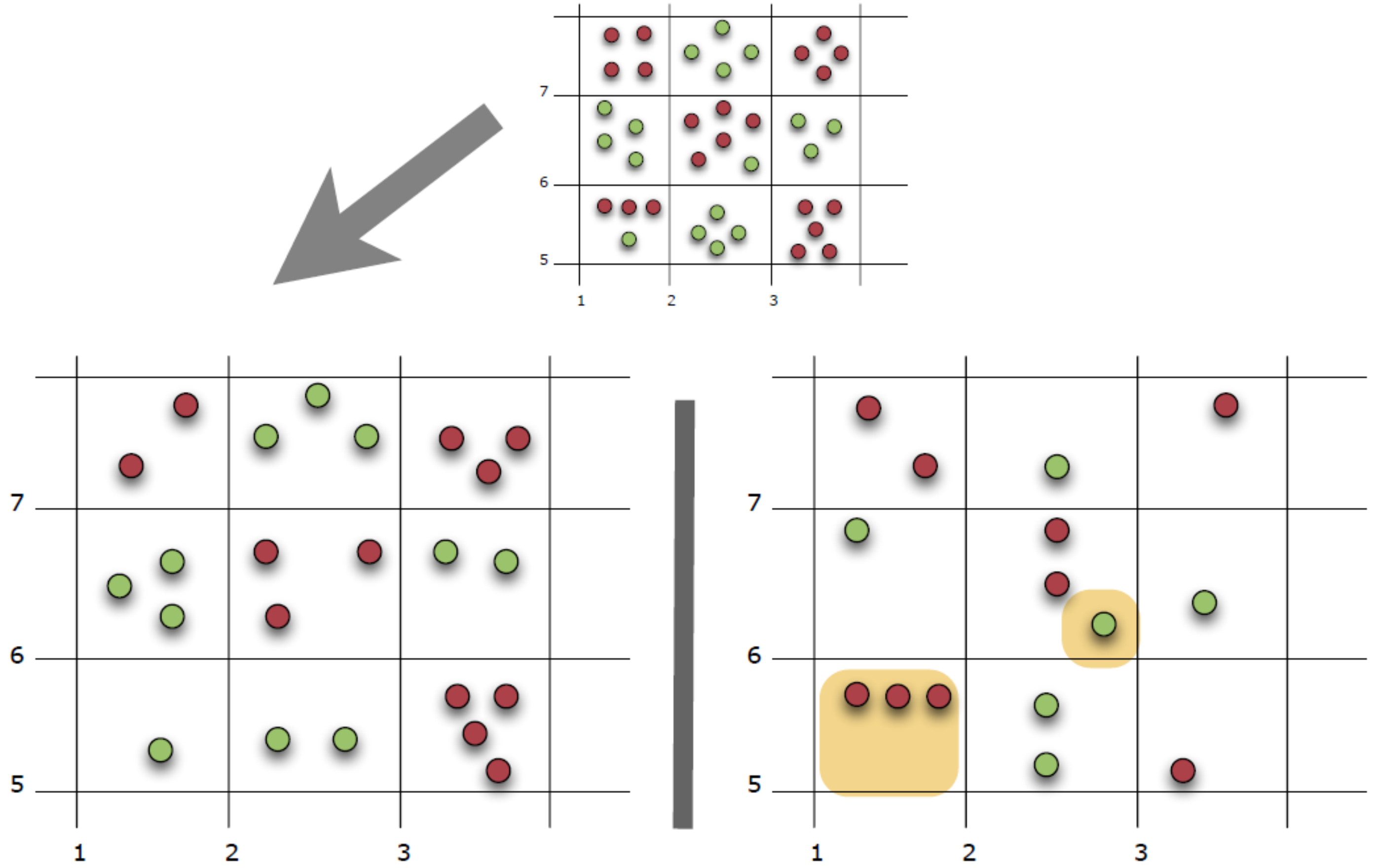


Figure 15. Example on various combination rules.

Bootstrap Aggregating: Bagging

- Out of bag error (OOB Error)
  - ✓ Use the training instances that are not sampled for validation

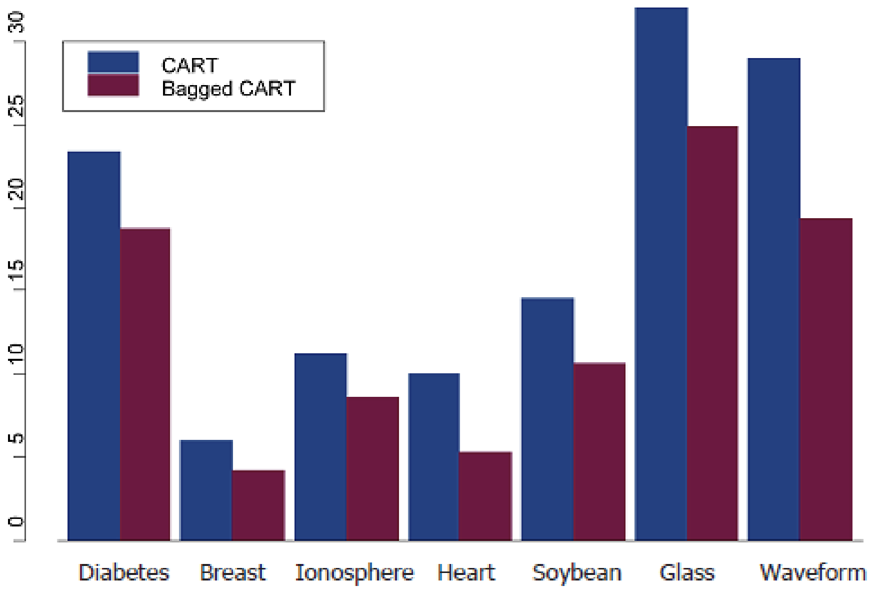


http://www.slideshare.net/oxdata/jan-vitek-distributedrandomforest522013

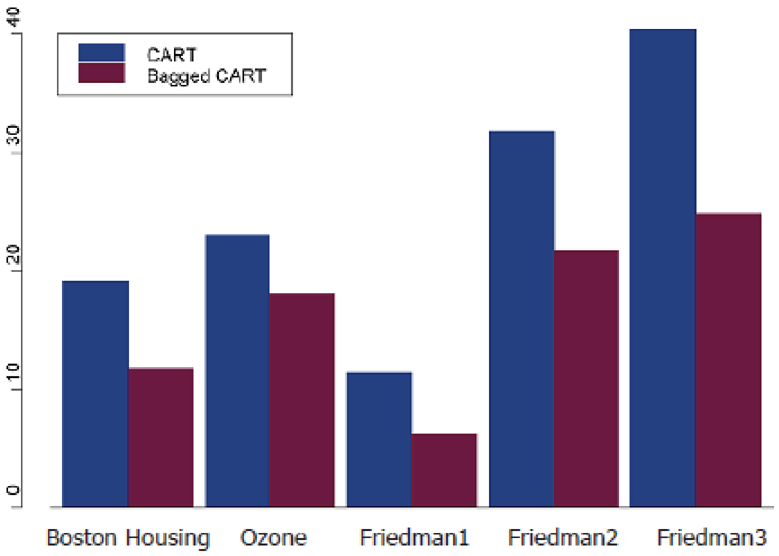
Bootstrap Aggregating: Bagging

- Bagged Trees vs. Single Tree

Classification



Regression




Q & A