An Introduction to Machine Learning

1 Keywords

- Accuracy
- Average Accuracy

$$\text{AA} = \frac{1}{C}\sum_c \frac{\sum_n \mathbf{1}(\widehat{y}_n = c \text{ and } y_n = c)}{\sum_n \mathbf{1}(y_n = c)} \tag{1}$$

- Average Precision

$$\text{AP} = \frac{1}{C}\sum_c \frac{\sum_n \mathbf{1}(\widehat{y}_n = c \text{ and } y_n = c)}{\sum_n \mathbf{1}(\widehat{y}_n = c)} \tag{2}$$

- Classification
- Confusion Matrix

$$C_{ij} = \sum_n \mathbf{1}(\widehat{y}_n = i)\mathbf{1}(y_n = j) \tag{3}$$

  This matrix can be used to compute `MA`, `OA`, `AA`, `AP`.

- Cross-Validation
- Dimension of the Feature Space
- Minimal Accuracy

$$\text{MA} = \min_c \frac{\sum_n \mathbf{1}(\widehat{y}_n = c \text{ and } y_n = c)}{\sum_n \mathbf{1}(y_n = c)} \tag{4}$$

- Overall Accuracy

$$\text{OA} = \frac{1}{N}\sum_n \mathbf{1}(\widehat{y}_n = y_n) \tag{5}$$

  where $X = [x_n]_n$ is the data matrix in the testing set, and x_n are different lines. $\widehat{y_n}$ is the class predicted by a given predictor. y_n is the true class for this sample. N is the number of lines of X.

- Testing Set
- Training Set
- Validation Set

2 Mathematical notations

- y_n is the true class to which sample n belongs to.
- $\widehat{y}_n$ is the predicted class for sample n.
- $\mathbf{1}(\text{"Statement"})$ is equal to 0 if "Statement" is wrong and 1 if "Statement" is correct.