==

ALGORITHM: BAYES CLASSIFIER

Let the training dataset $\mathbf{D}$ consist of n points $\mathbf{x}_i$ in a d-dimensional space, and let y_i denote the class for each point, with $y_i \in \{c_1, c_2, \ldots, c_k\}$.

The Bayes classifier estimates the posterior probability $P(c_i|\mathbf{x})$ for each class c_i, and chooses the class that has the largest probability. The predicted class for $\mathbf{x}$ is given as

$$\hat{y} = \arg\max_{c_i}\{P(c_i|\mathbf{x})\}$$

According to the Bayes theorem, we have

$$P(c_i|\mathbf{x}) = \frac{P(\mathbf{x}|c_i) \cdot P(c_i)}{P(\mathbf{x})}$$

Because $P(\mathbf{x})$ is fixed for a given point, Bayes rule can be rewritten as

$$\hat{y} = \arg\max_{c_i}\{P(c_i|\mathbf{x})\} = \arg\max_{c_i}\left\{\frac{P(\mathbf{x}|c_i)P(c_i)}{P(\mathbf{x})}\right\} = \arg\max_{c_i}\{P(\mathbf{x}|c_i)P(c_i)\}$$

ESTIMATING THE PRIOR PROBABILITY

Let $\mathbf{D}_i$ denote the subset of points in $\mathbf{D}$ that are labeled with class c_i:

$$\mathbf{D}_i = \{\mathbf{x}_j \in \mathbf{D} \mid \mathbf{x}_j \text{ has class } y_j = c_i\}$$

Let the size of the dataset $\mathbf{D}$ be given as $|\mathbf{D}| = n$, and let the size of each class-specific subset $\mathbf{D}_i$ be given as $|\mathbf{D}_i| = n_i$.

The prior probability for class c_i can be estimated as follows:

$$\hat{P}(c_i) = \frac{n_i}{n}$$

ESTIMATING THE LIKELIHOOD: NUMERIC ATTRIBUTES, PARAMETRIC APPROACH

To estimate the likelihood $P(\mathbf{x}|c_i)$, we have to estimate the joint probability of $\mathbf{x}$ across all the d dimensions, i.e., we have to estimate $P\big(\mathbf{x} = (x_1, x_2, \dots, x_d)|c_i\big)$.

In the parametric approach we assume that each class c_i is normally distributed, and we use the estimated mean $\hat{\boldsymbol{\mu}}_i$ and covariance matrix $\widehat{\boldsymbol{\Sigma}}_i$ to compute the probability density at $\mathbf{x}$

$$\hat{f}_i(\mathbf{x}) = \hat{f}(\mathbf{x}|\hat{\boldsymbol{\mu}}_i, \widehat{\boldsymbol{\Sigma}}_i) = \frac{1}{(\sqrt{2\pi})^d \sqrt{|\widehat{\boldsymbol{\Sigma}}_i|}} \exp\left\{ -\frac{(\mathbf{x} - \hat{\boldsymbol{\mu}}_i)^T \widehat{\boldsymbol{\Sigma}}_i^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_i)}{2} \right\}$$

The posterior probability is then given as

$$P(c_i|\mathbf{x}) = \frac{\hat{f}_i(\mathbf{x}) P(c_i)}{\sum_{j=1}^{k} \hat{f}_j(\mathbf{x}) P(c_j)}$$

The predicted class for $\mathbf{x}$ is:

HOW TO CLASSIFY A TEST POINT X?

To classify a numeric test point $\boldsymbol{x}$, the Bayes classifier estimates the parameters via the sample ***mean*** and ***sample covariance matrix.***

The *sample mean* for the class ci can be estimated as follows:

$$\hat{\mu}_i = \frac{1}{n_i} \sum_{x_j \in D_i} x_j$$

The *sample covariance matrix* for each class can be estimated as follows:

$$\hat{\Sigma}_i = \frac{1}{n_i} \boldsymbol{Z}_i^T \boldsymbol{Z}_i$$

where $\boldsymbol{Z}_i$ is the centered data matrix for class c_i, given as $\boldsymbol{Z}_i = \boldsymbol{D}_i - 1 \bullet \mu_i^T$. These values can be used to estimate the probability density as $f_i(x) = f\big(x \mid \hat{\mu}_i, \hat{\Sigma}_i\big)$

==

ALGORITHM: Bayes Classifier

BayesClassifier$\left(D = \left(x_j, y_j\right)_{j=1}^{n}\right)$:

for $i = 1, \dots, k$ **do:**

$D_i \leftarrow \{x_j | y_j = c_i, j = 1, \dots, n\}$ // Class-specific subsets

$n_i \leftarrow |D_i|$ // Cardinality

$\hat{p}(c_i) \leftarrow \frac{n_i}{n}$ // Prior probability

$\hat{u}_i \leftarrow \frac{1}{n_i} \sum_{x_j \in D_i} x_j$ // Mean

$Z_i = D_i - 1_{n_i} \cdot \hat{\mu}_i^T$ // Centered data

$\hat{\Sigma}_i \leftarrow \frac{1}{n_i} \boldsymbol{Z}_i^T \boldsymbol{Z}_i$ // Covariance matrix

return $\hat{p}(c_i)$**,** $\hat{u}_i$ and $\hat{\Sigma}_i$ for all $\boldsymbol{i = 1, \dots, k}$

TESTING ($\boldsymbol{x}$ and $\hat{p}(c_i), \hat{u}_i, \hat{\Sigma}_i$, for all $i \in [1, k]$):

$\hat{y}_i \leftarrow argmax_{c_i}\{f\left(x | \hat{u}_i, \hat{\Sigma}_i\right) \cdot \hat{p}(c_i)\}$

return $\hat{y}$

References

Mohammed, Z. J. & Wagner, M., 2014. *Data Mining and Analysis: Fundamental Concepts and Algorithms.* New York: Cambridge University Press.

Tan, P.-N., Steinbach, M., Karpatne, A. & Kumar, V., 2019. *Introduction to Data Mining.* 2nd ed. New York: Pearson Education.