==

ALGORITHM: NAÏVE BAYES CLASSIFIER FOR NUMERICAL ATTRIBUTES

The naive Bayes approach makes the simple assumption that all the attributes are independent, which implies that the likelihood can be decomposed into a product of dimension-wise probabilities:

$$P(\mathbf{x}|c_i) = P(x_1, x_2, \ldots, x_d|c_i) = \prod_{j=1}^{d} P(x_j|c_i)$$

The likelihood for class c_i, for dimension X_j, is given as

$$P(x_j|c_i) \propto f(x_j|\hat{\mu}_{ij}, \hat{\sigma}_{ij}^2) = \frac{1}{\sqrt{2\pi}\hat{\sigma}_{ij}} \exp\left\{-\frac{(x_j - \hat{\mu}_{ij})^2}{2\hat{\sigma}_{ij}^2}\right\}$$

where $\hat{\mu}_{ij}$ and $\hat{\sigma}_{ij}^2$ denote the estimated mean and variance for attribute X_j, for class c_i.

The naive assumption corresponds to setting all the covariances to zero in $\widehat{\mathbf{\Sigma}}_i$, that is,

$$\mathbf{\Sigma}_i = \begin{pmatrix} \sigma_{i1}^2 & 0 & \ldots & 0 \\ 0 & \sigma_{i2}^2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \ldots & \sigma_{id}^2 \end{pmatrix}$$

The naive Bayes classifier thus uses the sample mean $\hat{\boldsymbol{\mu}}_i = (\hat{\mu}_{i1}, \ldots, \hat{\mu}_{id})^T$ and a *diagonal* sample covariance matrix $\widehat{\mathbf{\Sigma}}_i = diag(\sigma_{i1}^2, \ldots, \sigma_{id}^2)$ for each class c_i. In total $2d$ parameters have to be estimated, corresponding to the sample mean and sample variance for each dimension X_j.

==

ALGORITHM: Naïve Bayes Classifier

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

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

$D_i \leftarrow \{x_j | y_j = c_i, j = 1, \ldots, n\}$ // 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

$Z_i = D_i - 1 \cdot \hat{\mu}_i^T$ // Centered data for class c_i

for $j = 1, \ldots, d$ **do**:

$\sigma_{ij}^2 \leftarrow \frac{1}{n_i} Z_{ij}^T Z_{ij}$ // Variance

$\hat{\sigma}_i = (\hat{\sigma}_{i1}^2, \ldots, \hat{\sigma}_{id}^2)^T$ // class specific attribute variances

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

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

$\hat{y}_i \leftarrow argmax_{c_i}\left\{\hat{P}(c_i) \prod_{j=1}^{d} f\left(x_j | \hat{\mu}_{ij}, \hat{\sigma}_{ij}^2\right)\right\}$

return $\hat{y}$

ALGORITHM DESCRIPTION

- Given the input dataset **D,** the method estimates the ***prior probability*** and the ***mean*** for each class.
- Next, it computes the variances σ_{ij}^2 for each of the attributes X_j, with all the d variances for class c_i stored in the vector σ_i. The variance for attribute X_j is obtained by first centering the data for class $\boldsymbol{D}_i$ via $\boldsymbol{Z}_i = \boldsymbol{D}_i - 1 \bullet \mu_i^T$.
- We denote by $\boldsymbol{Z}_{ij}$ the centered data for class c_i corresponding to attribute X_j. The variance is then given as $\sigma = \frac{1}{n_i} \boldsymbol{Z}_{ij}^T \boldsymbol{Z}_{ij}$.

CHARACTERISTICS OF THE NAÏVE BAYES CLASSIFIER

i. Naïve Bayes classifiers are probabilistic classification models that are able to quantify the uncertainty in predictions by providing posterior probability estimates.
ii. They are also generative classification models as they treat the target class as the causative factor for generating the data instances.
iii. By using the naïve assumption, they can easily compute class conditional probabilities even in high dimensional settings, provided that the attributes are conditionally independent of each other given the label. Suitable for diverse applications problems, such as text classification
iv. Naïve Bayes classifiers can handle missing values in the training set by ignoring the missing values of every attribute while computing its conditional probability estimates
v. Naïve Bayes classifiers are robust to irrelevant attributes.
vi. Correlated attributes can degrade the performance of the naïve Bayes classifiers because the naïve Bayes assumption of conditional independence no longer holds for such attributes.

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.