

Clasificación Bayesiana Ingenua:

1.1 Objetivo:

En este proyecto se busca que el estudiante entrene un modelo de clasificación de Aprendizaje Automático para predecir la clase de una muestra empleando en teorema de Bayes [Bishop, 2007].

1.2 Descripción Modelo:

Consideramos un conjunto de muestras o características $\mathbf{X} = \{\mathbf{x}_n\}_{n=1}^N$, con $\mathbf{x}_i \in \mathbb{R}^P$, y el conjunto de etiquetas $\mathbf{y} \in \mathbb{Z}^N$ con elementos enteros que representan la clase a la que pertenece cada muestra. Los datos se describen como

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1P} \\ x_{21} & x_{22} & \dots & x_{2P} \\ \vdots & & \ddots & \vdots \\ x_{N1} & x_{N2} & \dots & x_{NP} \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}.$$

Y la probabilidad de que una muestra o vector de características haya sido generado por una clase en particular (verosimilitud) se describe con una distribución normal

$$p(\mathbf{x} | y = c) = \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) = \frac{1}{(2\pi)^{P/2} |\boldsymbol{\Sigma}_c|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_c)^\top \boldsymbol{\Sigma}_c^{-1} (\mathbf{x} - \boldsymbol{\mu}_c) \right].$$

Empleando el teorema de Bayes se encuentra la probabilidad de las clases o etiquetas dada la muestra

$$p(y = c | \mathbf{x}) = \frac{p(\mathbf{x} | y = c)p(y = c)}{p(\mathbf{x})} \propto p(\mathbf{x} | y = c)p(y = c).$$

Bases de datos

- Se puede emplear la base de datos que utilizaron en el primero proyecto de “selección de características” o “Eigen Faces”.

Procedimiento

Para realizar la clasificación Bayesiana pueden emplear la librería scikit-learn. El procedimiento varía dependiendo de si el proyecto anterior fue “selección de características” o “Eigen Faces”

1.2.1 Selección de Características

- Realizar un análisis de cada variables de la base de datos por medio de Análisis exploratorio de datos (EDA).
- Particionar el conjunto de muestras en Entrenamiento (train) y Validación (test).
- Entrenar el clasificador Bayesiano con todo el conjunto de características y evaluarlo.
- Entrenar el clasificador Bayesiano empleando las características seleccionadas y evaluarlo.
- Comparar y discutir los resultados obtenidos.

1.2.2 Eigen-Faces

- Realizar un análisis de cada variables de la base de datos por medio de Análisis exploratorio de datos (EDA).
- Particionar el conjunto de muestras en Entrenamiento (train) y Validación (test).
- Entrenar el clasificador Bayesiano con todo el conjunto de características y evaluarlo.
- Proyectar el conjunto de entrenamiento sobre las 3 primeras componentes principales.
- Entrenar el clasificador Bayesiano empleando el conjunto de entrenamiento proyectado y evaluarlo.
- Comparar y discutir los resultados obtenidos.

Referencias

[Bishop, 2007] Bishop, C. M. (2007). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 1 edition.