

Práctica 4. Naïve Bayes

DRA. CONSUELO VARINIA GARCÍA MENDOZA

Especificaciones

- Para esta práctica se utilizará los dataset *iris.csv* y *emails.csv*
 - En el dataset *iris.csv*
 - Las primeras 4 columnas son las características de las instancias
 - La última columna es la clase
 - En el dataset *emails.csv*
 - La primera columna indica el id del correo
 - La última columna indica si el correo es spam o no
 - El resto de las columnas (3,000) son las palabras más comunes en todos los correos

Especificaciones

- Con los dataset iris.csv y email.csv realiza lo siguiente:
 - Carga el dataset
 - Crea un conjunto de entrenamiento del 70% de las instancias y el conjunto de pruebas con el 30% (*set random_state = 0*)
 - Utiliza el conjunto de entrenamiento para crear un conjunto de validación con 3 pliegues
 - Utiliza Naive Bayes con distribución normal para entrenar y probar el modelo en cada pliegue
 - Determina el accuracy de cada pliegue
 - Determina el accuracy promedio de los 3 pliegues
 - Utiliza Naïve Bayes con distribución multinomial para entrenar y probar el modelo en cada pliegue
 - Determina el accuracy de cada pliegue
 - Determina el accuracy promedio de los 3 pliegues
 - Selecciona el clasificador que obtenga el mejor accuracy promedio y utilízalo para entrenar el 70% de los datos (conjunto de entrenamiento) y predecir las clases del 30% de los datos (conjunto de pruebas)

Salida el programa

- Para los datasets iris.csv y emails.csv
 - Naive Bayes y distribución normal
 - Accuracy por pliegue
 - Accuracy promedio
 - Naive Bayes y distribución normal
 - Accuracy por pliegue
 - Accuracy promedio
 - Reporte de clasificación
 - Reporte de clasificación con el mejor clasificador(conjunto de prueba 70% y conjunto de entrenamiento 30%)
 - Matriz de confusión con `confusion_matrix` y `ConfusionMatrixDisplay`

Funciones

`sklearn.naive_bayes.GaussianNB`

`sklearn.naive_bayes.MultinomialNB`

`sklearn.metrics.accuracy_score`

`sklearn.metrics.classification_report`

`sklearn.metrics.confusion_matrix`

`sklearn.metrics.ConfusionMatrixDisplay`