



Escuela Superior de Cómputo
Machine Learning
Profesora: Consuelo Varinia García Mendoza
Práctica 3. Naïve Bayes



1. 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
2. Para cada dataset
 - Mezcla los datos (`random_state=0`) y crea un conjunto de entrenamiento del 70% y 30% de prueba
 - Con el 70% de entrenamiento genera conjuntos de validación con el método de validación cruzada para $k=5$. Genera la información de la Tabla 1 utilizando las bibliotecas:
 - `sklearn.model_selection.KFold`
 - `sklearn.naive_bayes.GaussianNB`
 - `sklearn.naive_bayes.MultinomialNB`
 - `sklearn.metrics.accuracy_score`
 - Selecciona las configuraciones que logran los mejores accuracy promedio para las pruebas finales. Genera la matriz de confusión y el reporte de clasificación utilizando las bibliotecas:
 - `sklearn.metrics.classification_report`
 - `sklearn.metrics.confusion_matrix`
 - `sklearn.metrics.ConfusionMatrixDisplay`
3. La salida del programa serán las tablas 1 y 2, las matrices de confusión y los reportes de clasificación (matrices y reportes sólo de las pruebas finales)

Tabla 1. Resultados de la validación cruzada

Dataset	No. Pliegues	Distribución	Pliegue	Accuracy
iris.csv	5	Normal	1	
			2	
			3	
			4	
			5	
			Promedio	
		Multinomial	1	
			2	
			3	
			4	
			5	
			Promedio	
emails.csv	5	Normal	1	
			2	
			3	
			4	
			5	
			Promedio	
		Multinomial	1	
			2	
			3	
			4	
			5	
			Promedio	

Tabla 2. Resultados de las pruebas finales

Dataset	Distribución	Accuracy
iris.csv		
emails.csv		