

Inteligencia Artificial 2026

Lab 02

16.febrero.2026

En este laboratorio vamos a implementar varios métodos de aprendizaje supervisado y aplicarlos a conjuntos de datos.

1. Para los siguientes datos: Construir el clasificador bayesiano óptimo. Indicar la regla de clasificación. Hallar el error de este

	$X = 1$	$X = 2$	$X = 3$	$X = 4$
$Y = 0$	0.09	0.16	0.17	0.05
$Y = 1$	0.15	0.07	0.10	0.21

clasificador.

2. Construir un clasificador bayesiano óptimo para dos distribuciones continuas con densidad exponencial, $f_0 \sim \text{Exp}(1)$ y $f_1 \sim \text{Exp}(3)$. Esto es, distribuciones con densidades condicionales

$$f_0(x) = f(x \mid Y = 0) = e^{-x}, \quad \text{y} \quad f_1(x) = f(x \mid Y = 1) = 3e^{-3x}.$$

En cada caso, construir la regla de clasificación, las regiones de clasificación y calcular el error del clasificador.

3. La siguiente tabla de datos muestra información de la decisión de jugar o no jugar golf, en función de varios factores como el tipo de clima, temperatura, humedad, o si hay viento, para 14 observaciones diferentes.

Obs.	Outlook	Temperature	Humidity	Windy	Play Golf?
0	Rainy	Hot	High	False	No
1	Rainy	Hot	High	True	No
2	Overcast	Hot	High	False	Yes
3	Sunny	Mild	High	False	Yes
4	Sunny	Cool	Normal	False	Yes
5	Sunny	Cool	Normal	True	No
6	Overcast	Cool	Normal	True	Yes
7	Rainy	Mild	High	False	No
8	Rainy	Cool	Normal	False	Yes
9	Sunny	Mild	Normal	False	Yes
10	Rainy	Mild	Normal	True	Yes
11	Overcast	Mild	High	True	Yes
12	Overcast	Hot	Normal	False	Yes
13	Sunny	Mild	High	True	No

Las variables son: $X_1 = \text{Outlook}$, $X_2 = \text{Temperature}$, $X_3 = \text{Humidity}$ y $X_4 = \text{Windy}$. Con base en la evidencia de la tabla, construya dos de los siguientes clasificadores (a) KNN, (b) Naïve Bayes, (c) Regresión logística para determinar si se debe o no jugar golf en los siguientes casos:

- $\mathbf{x} = (\text{Rainy}, \text{Hot}, \text{High}, \text{False})$.
- $\mathbf{x} = (\text{Sunny}, \text{Hot}, \text{Normal}, \text{False})$.

Compare el desempeño de sus dos clasificadores e indique con cuál se obtienen mejores resultados. Utilice métricas como Accuracy, Precision, Recall, F_1 Score, ROC AUC, para validar su argumento.

Responda la siguiente pregunta: ¿Cómo construir un clasificador bueno, con tan pocos datos?

4. Construir un clasificador binario para el siguiente conjunto de datos de Spam / No Spam. https://www.kaggle.com/datasets/alexandercarreio/es-spamham-dataset-para-deteccion-de-spam?select=spam_ham_es_dataset_v1.csv.

Usted puede elegir cualquiera de los algoritmos vistos en clase para construir su clasificador, pero el objetivo es utilizar aquel con mejor desempeño (menor error) posible. Utilice métricas como Accuracy, Precision, Recall, F_1 Score, ROC AUC, Binary Crossentropy, para validar el desempeño de su clasificador.

Para ello, cuando su algoritmo así lo requiera, deberá construir un *embedding* adecuado de sus textos a un vector numérico. Sugerencias:

- No olvide hacer un uso adecuado de el conjunto de entrenamiento, el conjunto de validación y el conjunto de prueba.
- Uso de técnicas de validación cruzada para una buena medición del desempeño.
- Se sugiere el uso de técnicas de procesamiento de lenguaje natural (NLP), para ello, investigue en qué consiste y para qué sirven los siguientes conceptos, y aplíquelas en su clasificador si lo considera conveniente:
 - remoción de *stopwords*
 - uso de *lemmatization*
 - Uso de vectorizaciones: *bag-of-words*, *TF-IDF*, *Doc2Vec*.