

## Tarea 2

Patrones y aprendizaje - Grupo 7093

Entrega: Viernes 11 de diciembre

Instrucciones: Respondan las siguientes preguntas en un notebook de Jupyter y súbanlo al classroom.

1. Cada una de las líneas en el documento `tarea2_docs.txt` lo vamos a considerar como un documento. Realiza la limpieza y preprocesamiento necesarios para contestar las siguientes preguntas:
  - ¿En qué tópicos se pueden dividir?
  - ¿Qué documentos hablan de México y su cultura?
2. Descarga el dataset de datos de sonar de la siguiente liga:  
<https://archive.ics.uci.edu/ml/machine-learning-databases/undocumented/connectionist-bench/sonar/sonar.all-data>  
mismo que deberás explorar y entender lo que hacen sus atributos.  
Lleva a cabo una clasificación utilizando la regresión logística de los datos. Luego efectúa una reducción de dimensión y haz tu clasificación otra vez. La comparación entre las clasificaciones la harás con las métricas que tú elijas.  
  
¿Qué observas y por qué sucede esto?
3. Diseña un experimento para determinar para qué umbral es posible tener un conjunto de datos no balanceado.
4. Reproduce la gráfica de la información mutua del seno contra él mismo considerando un corrimiento de 50 posiciones.
5. El dataset de `seatbelt.csv` representa una serie de datos temporales de diversos atributos que hacen referencia a la mortalidad asociada a los accidentes de tránsito en Gran Bretaña en el periodo de 1969 a 1984. La legislación para utilizar el cinturón de seguridad de manera obligatoria fue introducida el 31 de enero de 1983. Ajusta un modelo lineal generalizado para determinar la probabilidad de morir en un accidente de tránsito y responde las siguientes pregunta:
  - ¿Volver obligatorio el cinturón de seguridad disminuyó la probabilidad de morir en algún accidente de tránsito?

- ¿Qué otra conclusión puedes generar a partir del modelo que ajustaste? (texto libre)

Se espera que lleves a cabo la normalización o estandarización del dataset, la transformación de las variables que consideres pertinente y que propongas qué variables tienen más o menos importancia en la probabilidad de morir en un accidente de tránsito.