

Lista de Ejercicios; AM.

Lista 2.

1. Ejercicios; 2.1, 2.2, 2.3, 2.12.
2. Genere 500 muestras a partir de dos normales multivariadas N_1, N_2 con parametros $\mu_1 = [1, 5 \ 1, 5]^T, \mu_2 = [5, 0 \ 5, 0]^T$ y $\Sigma_1 = \Sigma_2 = \begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix}$ de tal suerte que se generen, primero tres muestras con N_1 y luego dos muestras con N_2 . Con los datos generados aplique el algoritmo *AlgoritmoEM_2Gaussianas.py* en los siguientes dos casos ;
 - a) Con los datos ordenados, aplique el algoritmo anterior, hasta la iteración $n = 100$, con datos iniciales $\mu_1(0) = [1, 05, 1, 15]^T, \mu_2(0) = [1, 35, 1, 65]^T, \sigma_1^2 = 0, 3, \sigma_2^2 = 0, 04$ y $P = 0, 65$.
 - b) Repita el inciso a), pero esta vez mezclando las muestras simuladas.
 - c) Repita el inciso a), pero esta vez cambie los datos iniciales.
Compare los resultados obtenidos en a), b) y c).
3. Implemente el algoritmo perceptrón de la página 94.
4. En el código *Pocket_Algorithm.py* identifique los pasos de Algoritmo Pocket.
Hint: Apoyese de las páginas 94 y 101.
A continuación trabajaremos con los datasets
 - a) Boston house-prices dataset.
 - b) Diabetes dataset.
 - c) Iris dataset.
 - d) Wisconsin breast cancer dataset.
 - e) Wine dataset.
de la biblioteca *Scikit-learn*
5. En un solo dataset, registre los nombres de los datasets anteriores, las observaciones de cada dataset, las dimensiones de la matriz de características, los tipos de datos de las característica y el tipo de dato de las etiquetas.
Ejemplo.

Dataset name	Observations	Dimensions	Features	Targets
Boston house-prices	506	13	real, positive	real 5.–50.

6. Cree un arreglo de numpy, llamado *data* como el de la siguiente figura

```
array([[ 5., nan,  8.],
       [ 9.,  3.,  5.],
       [ 8.,  6.,  4.],
       [nan,  5.,  2.],
       [ 2.,  3.,  9.],
       [nan,  8.,  7.],
       [ 1., nan,  5.]])
```

del paquete `sklearn.impute` importe la función `SimpleImputer` y utilízela para llenar los datos faltantes utilizando la estrategia `'mean'`.

7. Construya un modelo de clasificación SVC para resolver la tarea de clasificar el dataset *Wine dataset*, utilizando a) el 30 % de los datos para entrenamiento, y b) el 70 % de los datos para entrenamiento. Utilizando la métrica *accuracy_score*. ¿Se puede observar el fenómeno de overfitting?. Explique.
8. Construya un modelo de clasificación SVR para resolver la tarea de regresión en el *Diabetes dataset*, utilice el 35 % de los datos para entrenar su modelo y la métrica *mean_squared_error*.
9. Explique el método con el cual se *crece* un árbol de decisión para una tarea de clasificación y para una tarea de regresión
10. Resuelva los problemas 7 y 8, utilizando modelos de árboles de decisión.
11. Los archivos *Data0.csv* y *Data1.csv* contienen los datos de dos distintas series de tiempo, de los precios de dos activos. Construya modelos de regresión utilizando la máquina de vectores de soporte, árboles de decisión y bosque aleatorio para resolver la tarea de regresión de ambas series. Compare sus resultados.