

Seminario de Aplicaciones Actuariales

Seminario de Estadística I

Aplicaciones de Ciencia de Datos con Python

Profesor: Dr. Arrigo Coen Coria

Ayudante: Act. Miriam Colín

Tarea 2: Algoritmos de clasificación

Instrucciones:

- La entrega será el **viernes 23 de julio**. Puede ser de manera individual o en equipos de a lo más 3 alumnos.
- Las preguntas 1-10 se entregarán en un pdf con el nombre:
T1_ApPatNom1_ ApPatNom2_ ApPatNom3_1_10
- Cada una de las preguntas 11-16 se entregarán en un jupyter notebook con los nombres:
T1_ApPatNom1_ ApPatNom2_ ApPatNom3_11,
T1_ApPatNom1_ ApPatNom2_ ApPatNom3_12,
T1_ApPatNom1_ ApPatNom2_ ApPatNom3_13,
T1_ApPatNom1_ ApPatNom2_ ApPatNom3_14,
T1_ApPatNom1_ ApPatNom2_ ApPatNom3_15,
T1_ApPatNom1_ ApPatNom2_ ApPatNom3_16, respectivamente.
- Responde las siguientes preguntas y realiza lo que se pide.

1. Responde las siguientes preguntas:

- 1.1. ¿Cuál es la idea fundamental de SVM?
- 1.2. ¿Qué es un vector de soporte?
- 1.3. ¿Por qué es importante escalar los insumos de SVM?
- 1.4. ¿Cómo opera el margen suave?
- 1.5. ¿Cómo usar regularización en SVM?

2. Utiliza SVM para clasificar la salud de los embarazos de la base de datos <https://www.kaggle.com/andrewmvd/fetal-health-classification>.

3. 9. Train an SVM classifier on the MNIST dataset. Since SVM classifiers are binary

classifiers, you will need to use one-versus-the-rest to classify all 10 digits. You may want to tune the hyperparameters using small validation sets to speed up the process. What accuracy can you reach?

1. 10. Train an SVM regressor on the California housing dataset. son los pasos para realizar un ajuste de ML con Python?
2. ¿En qué consisten los pasos para realizar un ajuste de ML con Python?

3. Describe para qué se utilizan las siguientes librerías de Python:
 - Math
 - Matplotlib
 - Numpy
 - Pandas
 - Seaborn
 - Sklearn
4. ¿Cuáles son las etapas de la metodología CRISP-DM?
5. Eres el encargado de un nuevo proyecto para mejorar el algoritmo de recomendación de canciones de *Spotify* y tu jefe te pide que apliques la metodología CRISP-DM. Describe, a grandes rasgos, cuales son los posibles desafíos que te enfrentarías en cada una de las etapas de CRISP-DM.
6. Con la base de datos del archivo *T1_cars* realizar un análisis completo de los datos y una regresión lineal simple.
7. Con la base de datos *California Housing Prices* realizar un análisis completo de los datos y una regresión lineal múltiple.
//sklearn.datasets.fetch california housing.
8. Con la base de datos del archivo *T1_red_wine* realizar un análisis completo de los datos y una regresión *Ridge*.
9. Con la base de datos del archivo *T1_white_wine* realizar un análisis completo de los datos y una regresión *LASSO*.
10. Con la base de datos de "**House Prices**" de Kaggle definir un modelo de regresión ya sea Lineal, Ridge o Lasso para estimar los precios de venta. Subir el su predicción a Kaggle y obtener un score menor que 0.143.
11. Realizar el ejercicio del archivo *2_2-1 Linear Regression House Price Kaggle* con los nombres de su equipo. Las variables que van a utilizar se seleccionan de manera aleatoria cambiando la variable *team_members*.