**MACHINE LEARNING FOR GEOSCIENCES**

|  |  |
| --- | --- |
| **Student Name: Felipe Abraham Santana Román** | **Date: 19-03-2020** |

In problems 2, 3 and 6 replace for the one assigned to your name,

|  |  |
| --- | --- |
| Student |  |
| David | 2 |
| Felipe | 3 |
| Karla | 4 |
| Moises | 5 |
| Reynaldo | 6 |
| Yuri | 7 |

1. What is overfitting?

Tambien llamado como sobreajuste se refiere al hecho de que un modelo de aprendizaje este sobreentrenado en el que solo podra predecir los datos de la primera entrada, datos muy especificos, y no sera possible que pueda predecir datos de entradas posteriores.

1. What is the total error and the MSE for the model and the following instances?

|  |  |
| --- | --- |
|  |  |
| -1 | -6 |
| 0 | 7 |
| 3 | 3 |

1. Algebraically, find the best for the model and the instance

y-3x=b

sustituyendo

7-3(3)=b

7-9=b

-2=b

Sustituyendo el mejor modelo es

Y=3x-2

Write Python code for the following:

1. Define a character string and print it to the Python console.
2. Define a pandas DataFrame for the following table

|  |  |  |
| --- | --- | --- |
| Index | ML\_classification | Application percentage in Earth Sciences |
| 1 | Unsupervised learning | 40% |
| 2 | Supervised learning | 60% |

1. Create a DataFrame with columns and , provide numeric instances and create a scatterplot of it.
2. (3 pts.) Find the best model to predict permeability (column Perm) for the dataset in <https://raw.githubusercontent.com/GeostatsGuy/GeoDataSets/master/1WellPorPerm.csv>
3. Plot histogram of Porosity--Listo
4. Plot histogram of Permeability----Listo
5. Plot the scatterplot matrix using all 3 columns----Listo
6. Find the best regression model (linear, tree, …)
7. What is the numeric value of its training error?
8. What is the numeric value of the test set error?