

Proyecto #1

Prof. Cristian López Del Alamo ,
Prof. Vicente Machaca

Regresión

Objetivo

Implementar un algoritmo de regresión multivariada para predecir el área de terreno que se podría incendiar dependiendo de las características del terreno.

Actividades

1. Utilizar el *dataset* *Forest Fire*. Este *dataset* cuenta con 517 instancias, 12 atributos los cuales representan las características del suelo y un campo adicional denominado *area* que representa el área del terreno que podría incendiarse dependiendo de los atributos del suelo.
2. Separar el *dataset* en 70 % para entrenamiento, 20 % para validación y 10 % para *testing*. Es importante que cada uno de los grupos deban ser tomado de manera aleatoria.
3. Entrenar el modelo con el 70 % de los datos durante un número de épocas ingresada como hiper-parámetro.
4. En cada época, se entrena el modelo, es decir, se actualizan los pesos en la dirección de los gradientes y se calcula el error de entrenamiento, además, se valida el modelo con el 20 % de los datos, sin actualizar los pesos, y se calcula el error de validación. Almacenar en dos listas, los errores de entrenamiento y validación.
5. Graficar el error de entrenamiento y validación y diferenciar entre un *overfitting* y un *underfitting*.
6. Realizar pruebas con distintos valores de *alfa*
7. Pasar, por el modelo, el 10 % de los datos de *testing* y mostrar su error.

Cada equipo deberá subir un sólo documento con la siguiente estructura:

1. Introducción.
2. Explicación: Explicar el funcionamiento del modelo de regresión. El documento deberá ser escrito utilizando el formato *paper* el cual puede ser descargado de aquí: [ieee](#)
3. Experimentos: Deberá explicar qué y cómo se hicieron los experimentos, además de colocar sus resultados utilizando tablas y gráficos que soporten los experimentos.
4. Conclusiones: Deberá redactar las conclusiones de acuerdo a los resultados.

Nota: Toda copia parcial o total en el documento o código fuente invalida el trabajo de todo el equipo.

Finalmente, incluir en el documento, el link del github o el colab para verificar el código fuente. Así mismo, definir como *semilla*: `np.random.seed(0)`; para replicar los resultados.

Nota: El grupo debe estar formado por **máximo 3 personas**.