

Plan Formativo: Ciencia de Datos	Nivel de Dificultad
Módulo: Inferencia Estadística	Medio
Tema: Machine Learning	
Intención del aprendizaje o aprendizaje esperado:	
Elaborar un modelo predictivo de regresión lineal simple utilizando técnicas de validación cruzada y el lenguaje Python para resolver un problema.	
Ejercicios planteados	
Ejercicio: Suponga que le han pedido averiguar si existe una relación entre el número de veces que batean los jugadores de un equipo de béisbol y el número de <i>runs</i> que consigue. En caso de existir y de establecer un modelo, podría predecir el resultado del partido para los siguientes datos:	
<pre># Datos #===== equipos = ["Texas","Boston","Detroit","Kansas","St.","New_S.","New_Y.", "Milwaukee","Colorado","Houston","Baltimore","Los_An.","Chicago", "Cincinnati","Los_P.","Philadelphia","Chicago","Cleveland","Arizona",</pre>	

```
"Toronto", "Minnesota", "Florida", "Pittsburgh", "Oakland", "Tampa",  
    "Atlanta", "Washington", "San.F", "San.I", "Seattle"]  
bateos = [5659, 5710, 5563, 5672, 5532, 5600, 5518, 5447, 5544, 5598,  
    5585, 5436, 5549, 5612, 5513, 5579, 5502, 5509, 5421, 5559,  
    5487, 5508, 5421, 5452, 5436, 5528, 5441, 5486, 5417, 5421]  
  
runs = [855, 875, 787, 730, 762, 718, 867, 721, 735, 615, 708, 644, 654,  
    735, 667, 713, 654, 704, 731, 743, 619, 625, 610, 645, 707, 641,  
    624, 570, 593, 556]  
  
datos = pd.DataFrame({'equipos': equipos, 'bateos': bateos, 'runs':  
runs})
```

Para esto obtenga lo siguiente:

1. Obtenga la distribución de los datos
2. Separe sus datos de entrenamiento en 80/20.
3. Ajuste un modelo de regresión lineal.
4. Escriba la recta de regresión que obtuvo con sus estimaciones respectivas.
5. Obtenga la gráfica de los datos y su recta de regresión
6. Obtenga las métricas: MAE, MSE, RMSE.
7. Concluya con los datos obtenidos y comente la problemática que le han planteado al principio.

Probabilidad y estadística

Preguntas guía

- Algoritmos de regresión lineal simple
- Quién o qué es la variable respuesta

Recursos Bibliográficos:

[1] Modelos de regresión

<https://economipedia.com/definiciones/modelo-de-regresion.html>

[2] Librería Scikit-Learn de Python

<https://aprendeia.com/libreria-scikit-learn-de-python/>

[3] Métricas

[https://www.google.com/search?q=M%C3%A9tricas+de+evaluaci%C3%B3n+del+modelo+regresivo+\(MAE%2C+MSE%2C+RMSE\)&rlz=1C5CHFA_enHU941HU941&oq=M%C3%A9tricas++de++evaluaci%C3%B3n++del++modelo++regresivo++\(MAE%2C+MSE%2C+RMSE\)&aqs=chrome..69i57.454j0j7&sourceid=chrome&ie=UTF-8](https://www.google.com/search?q=M%C3%A9tricas+de+evaluaci%C3%B3n+del+modelo+regresivo+(MAE%2C+MSE%2C+RMSE)&rlz=1C5CHFA_enHU941HU941&oq=M%C3%A9tricas++de++evaluaci%C3%B3n++del++modelo++regresivo++(MAE%2C+MSE%2C+RMSE)&aqs=chrome..69i57.454j0j7&sourceid=chrome&ie=UTF-8)

[4] Modelos de Markov

<https://www.elsevier.es/es-revista-farmacia-hospitalaria-121-articulo-introduccion-utilizacion-modelos-markov-el-10017809>