

MACHINE LEARNING EN FINANZAS

Problem Set 1

INTRODUCCIÓN

La idea de este Problem Set es implementar un módulo con un Regresor compatible con el API de **scikit-learn**. Para ello deberá cumplir con los requisitos expuestos en la referencia:

<https://scikit-learn.org/stable/developers/develop.html#rolling-your-own-estimator>

REQUISITOS

1. El Regresor deberá llamarse `EstimadorOLS` y corresponderá con una regresión lineal. La fórmula de estimación de los parámetros corresponde a:

$$\hat{\beta} = (X'X)^{-1}X'y$$

donde X es la matriz con las series de las variables por columna (la primera columna contendrá todos 1 para estimar la constante), e y es el vector columna con los valores de la variable independiente.

Deberá aceptar un parámetro `fit_intercept` de tipo `bool` para indicar si agregar o no columna con 1 al dataset input, con default en `True`.

2. La rutina de testing del módulo deberá:
 - a. Para todos los casos relevantes eliminar la aleatoriedad entre simulaciones utilizando `seed 1234`.
 - b. Generar un Training Set con la función `sklearn.datasets.make_regression` para obtener 1000 observaciones de 10 variables con una constante de 5 y un desvío standard en el error de 10.
 - c. Realizar el test de validación utilizando `mlfin.printing.print_validation_results`
 - d. Ajustar el modelo a todo el Training Set e imprimir en pantalla el valor de la constante y los coeficientes estimados.
 - e. En pantalla debería verse el siguiente output:

```
Testeando <class '__main__.EstimadorOLS'>
Resultado de Cross-Validation:
Scores : [99.83%, 99.84%, 99.84%, 99.82%, 99.79%]
Mean    : 99.82%
Std Dev : 0.02%

Original coeffs:
[91.24180775 35.26752536 54.24529655 94.50110629 93.59602075 13.35326338
 99.01663251 67.11803665 55.4120574 88.83088879]

Fitted coeffs:
Constante: 5.156110
[91.60080555 34.94535293 54.28341038 94.20550105 94.00134648 13.25777365
 99.4225115 67.27709181 55.24055602 88.97083603]
```

ÉXITOS!