

MODELOS Y TECNOLOGÍAS DE LOS SISTEMAS DE INFORMACIÓN 2018-2019

Práctica 4:

1. Objetivo

Esta práctica consiste en aplicar algunos de los conceptos básicos vistos relativos a modelos de Aprendizaje Automático. En particular se trata de que el alumno sea capaz de crear un modelo de regression lineal múltiple empleando datos obtenidos de Yahoo Finance.

El problema consiste en crear un modelo capaz de predecir las rentabilidades futuras de un active basándose en las rentabilidades de los últimos 5 días del activo.

La practica se realizará en los mismo grupos que la practica anterior La nota de la práctica sera la misma para todos los componentes del equipo.

2. Recursos

El alumno debe emplear datos de cualquier valor de su elección de Yahoo Finance (<http://finance.yahoo.com/q/hp?s=%5EGSPC+Historical+Prices>) empleando las cotizaciones desde el primer día posterior al 1 de enero de 2019 (por ejemplo, 4 de enero) al 31 de marzo de 2019 o último día disponible antes de esa fecha.

Una vez obtenidas las cotizaciones, el alumno debe transformarlas en rentabilidades tomando la diferencia (a partir del segundo día disponible) de precios y dividiéndola por el precio anterior:

$$r_t = \frac{P_t - P_{t-1}}{P_{t-1}}, t = 2, 3, \dots, n$$

A partir de dichos rendimientos deberá construir una matriz de variables dependientes (r_t) e independientes teniendo en cuenta las rentabilidades de los cinco días anteriores además de un vector de unos ($1, r_{t-1}, \dots, r_{t-5}$). Evidentemente solo es posible construir dicha matriz a partir de las SEXTA observacion de rendimientos (o SEPTIMA de precios). Las matrices de las variables dependientes e independientes se muestran en el siguiente ejemplo con los colores rojo y naranja, respectivamente.

dia	P(t)	r(t)		r(t-1)	r(t-2)	r(t-3)	r(t-4)	r(t-5)
1	12							
2	13	0,083	1,000					
3	11	-0,154	1,000	0,083				
4	12	0,091	1,000	-0,154	0,083			
5	14	0,167	1,000	0,091	-0,154	0,083		
6	13	-0,071	1,000	0,167	0,091	-0,154	0,083	
7	11	-0,154	1,000	-0,071	0,167	0,091	-0,154	0,083
8	12	0,091	1,000	-0,154	-0,071	0,167	0,091	-0,154
9	13	0,083	1,000	0,091	-0,154	-0,071	0,167	0,091
10	15	0,154	1,000	0,083	0,091	-0,154	-0,071	0,167
....	11	-0,267	1,000	0,154	0,083	0,091	-0,154	-0,071

3. Tareas a realizar

Primera parte

1. A partir de dicha matriz el alumno debe construir un modelo de regresión lineal múltiple en python y presentar brevemente el valor de los parámetros encontrados y la función que representan (5 líneas). El alumno o grupo debe elegir el procedimiento de aprendizaje que considere más apropiado en función de la dimensionalidad del problema.
2. Una vez implementado dicho modelo el alumno deberá proporcionar dos columnas, una con el rendimiento real del activo y otra con el rendimiento predicho y comparar sus predicciones con las de un modelo que predice que SIEMPRE el rendimiento va a ser cero (paseo aleatorio)
3. Sobre los anteriores resultados el alumno deberá comentar brevemente:
 - a. Son los resultados buenos o malos? porque? (5 líneas)
 - b. Se le ocurre alguna forma de mejorar el modelo? (10 líneas)

Segunda parte

1. El alumno debe repetir la secciones 1. y 2. de la primera parte pero de manera iterada, es decir, considerando una ventana DESLIZANTE de UN DIA y considerando las treinta observaciones anteriores, así, por ejemplo, se pueden emplear los últimos 30 días para predecir el siguiente, mover un día la ventana (manteniendo los 30 últimos días) y volver a predecir, etc. En esta parte el alumno debe considerar TAMBIEN los datos disponibles en abril de 2019. Una vez hecho esto el alumno debe graficar el coeficiente del intercepto a lo largo de todas las predicciones y contestar a lo siguiente: cambia mucho? Porque si o porque no? que interpretación tiene esto? (5 líneas)
2. Suponga ahora que quiere emplear el modelo para invertir, suponga que la rentabilidad total es igual a la suma de las diferencias de rentabilidades entre las predichas y reales a lo largo de los días de predicción, cuál sería dicha rentabilidad? Es alta o baja? Podemos decir que el modelo ha funcionado? (5 líneas)

4. Formato

El alumno debe proporcionar EXCLUSIVAMENTE TRES archivos SIN COMPRIMIR (por separado)

- a) Un archivo con los datos empleados, en el formato que prefiera (csv, xls, etc.) pero que sea DIRECTAMENTE leído por el programa.
- b) Un archivo (o Notebook) en Python que deberá ejecutar TODAS las tareas requeridas así como presentar por pantalla los resultados solicitados. El código del programa deberá estar contenido en UN SOLO FICHERO, es decir, NO se podrán utilizar otras funciones salvo las incluidas en el propio entorno Python. Igualmente el programa deberá leer POR DEFECTO el fichero anterior, NO PUDIENDO EMPLEARSE otros directorios que requieran su creación previa.
- c) Una explicación y resumen de las tareas realizadas que NO PODRA EXCEDER una hoja por una sola cara EN PDF y en la que en el encabezado indicará: PRACTICA 4_APELLIDOS DE LOS ALUMNOS: MODELADO DE UNA APLICACION DE *PREDICCIÓN BURSÁTIL* BASADA EN APRENDIZAJE AUTOMÁTICO y los nombres y correos electrónicos de todos los integrantes del equipo.

5. Fecha

La práctica se entregará EXCLUSIVAMENTE a través del buzón habilitado en la plataforma que estará abierto hasta las **23.55 horas del lunes 22 de abril**. Cada equipo deberá enviar UN SOLO archivo (el envío lo realizará un miembro por equipo). La práctica podrá entregarse, por tanto, en cualquier momento anterior a dicha fecha y hora.

6. Criterios de valoración

La valoración de la práctica atenderá a los siguientes:

1. Adecuación a las normas aquí presentadas
2. Profundidad del análisis y originalidad en las soluciones propuestas
3. Corrección del código entregado