

MODELOS DE TECNOLOGÍAS PARA LOS SISTEMAS DE INFORMACIÓN 2022-2023

Práctica PEC 2 (OBLIGATORIA, ENTREGABLE):

1. Objetivo

Esta práctica consiste en aplicar los conceptos vistos en el tema relativos a la construcción de modelos lineales, *knn* y árboles en un problema de predicción dinámica en un contexto financiero. La idea es que el alumno sea capaz de implantar dicho modelo de Aprendizaje Automático en un problema financiero, siendo para ello necesario que el alumno implemente en Python dichos modelos, así como las métricas asociadas y otros análisis.

2. Planteamiento

La práctica plantea el uso y análisis de un modelo de Aprendizaje Automático y permite ilustrar cómo algunos modelos de Aprendizaje Automático pueden ser empleados en problemas de modelización dinámica, contextualizando conceptos vistos. Se espera también que el alumno desarrolle una visión crítica sobre dicho modelo.

Se trata de un problema de predicción dinámica. Para el que alumno deberá emplear datos del S&P500 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 2021 (por ejemplo, 4 de enero) al 28 de febrero de 2023 o último día de febrero disponible antes de esa fecha.

Una vez obtenidas las cotizaciones, el alumno debe transformarlas en rentabilidades (*rendimientos*) 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 observación de rendimientos (o SEPTIMA de precios), debiendo descartar las filas no completas.

3. Tareas a realizar y preguntas

A partir de dicha matriz el alumno debe construir un modelo de regresión lineal múltiple en Python del tipo $Y=a+b_1X_1+b_2X_2+...+b_5X_5$ y contestar a las preguntas mas abajo indicadas.

1. De cuantos parámetros consta el modelo? (0.5 puntos)

Explicar en una transparencia los resultados

2. Los parámetros correspondientes al tercer y cuarto regresor son iguales? (0.5 puntos)

Explicar en una transparencia los resultados

Una vez implementado dicho modelo el alumno deberá producir dos columnas, una con el rendimiento real del activo y otra con el rendimiento predicho y sobre ambas calcular tres métricas: el *mse*, el *mae* y el *mape*.

3. Las tres métricas, proporcionan los mismos resultados (el mismo valor)? (0.5 puntos)

Explicar en una transparencia los resultados

4. Suponga que comparamos las predicciones frente a un paseo aleatorio $r_t=0$ empleando el *mape*, las predicciones del modelo lineal son mejores? (0.5 puntos)

Explicar en una transparencia los resultados

Considere ahora un conjunto de entrenamiento que abarca desde el inicio hasta el 31 de diciembre de 2022 (o la fecha mas próxima disponible en 2022) y otro de prueba (*test*), que abarca desde el 1 de enero de 2023 (o la fecha mas próxima disponible de 2023) hasta el final de la serie.

5. Encuentre, sobre el conjunto de entrenamiento y empleando validación cruzada (10 conjuntos) el mejor modelo de vecinos próximos (algoritmo *knn*) empleando entre 1 y 15 vecinos. Con el modelo óptimo calcule el *ecm* sobre el conjunto de test y compárelo con un paseo aleatorio. Es mejor o peor que un paseo aleatorio? (2.25 puntos)

Explicar en tres transparencias los resultados

- 6.-Emplando el mismo procedimiento, métrica y *benchmark* con un árbol de regresión con una profundidad entre 1 y 10 niveles. Es mejor o peor que un paseo aleatorio? (2.25 puntos)

Explicar en tres transparencias los resultados

Para la siguiente pregunta el alumno debe considerar una ventana DESLIZANTE de UN DIA, considerando únicamente las doscientas observaciones anteriores, es decir, se emplean los últimos doscientos días para predecir el siguiente, se mueve un día la ventana y se vuelve a predecir, etc., El periodo predictivo comienza el 15 de febrero de 2023 (o primer día disponible después de esa fecha) y termina el 28 de febrero de 2023 (o el día más próximo disponible en febrero). Adicionalmente, suponemos que no hay ventas en corto, es decir, solo se puede invertir en caso de una predicción de precio superior al actual. Para knn y árboles se emplearán los hiper-parámetros óptimos encontrados anteriormente, es decir NO es preciso emplear validación cruzada.

7. La rentabilidad de cada uno de los modelos (lineal, knn y arboles) es positiva? (2.5 puntos)

Explicar en cuatro transparencias los resultados

8. Sugerir posibles mejoras al los procedimientos efectuados en 5. 6. y 7. (1 punto)

Explicar en dos transparencias los argumentos

A partir de los resultados anteriores, cada grupo deberá producir un conjunto de transparencias (obligatoriamente en formato *power point*) que sirvan de soporte para **una presentación de estilo pechakucha (20 transparencias, 6 min 40 seg. de presentación)**. La portada de la presentación debe indicar nombre de los alumnos así como "PRACTICA PEC2 ENTREGABLE: PREDICCION DINAMICA MEDIANTE MODLOS LINEALES, KNN Y ARBOLES". Adicionalmente se espera una breve introducción al problema aunque NO se espera la explicación detallada de los algoritmos en cada una de las secciones.

Para la presentación se espera que el alumno sea capaz de entender qué información es relevante y cual no, presentando exclusivamente ésta; en particular se aconseja el empleo de tablas sencillas y comprensibles, algunos gráficos igualmente simples y comprensibles y el empleo de poco texto y con un tamaño de letra suficiente.

5. Fecha

La práctica se entregará en un buzón disponible en *Blackboard* hasta las **22.00 horas del Domingo 26 de marzo**, dicha entrega debe contener **el código y la presentación en formato .ppt** (NO pdf u otros), no se admitirán archivos comprimidos. En la clase de prácticas del **27 de marzo**, a partir de las 15.00 horas y en el orden determinado en clase se presentarán en clase los trabajos de cada uno de los grupos. A las presentaciones solo puede acudir el grupo correspondiente.

6. Criterios de Valoración

- 1) Seguimiento de las normas
- 2) Originalidad
- 3) Corrección de los resultados
- 4) Presentación
- 5) Profundidad de los análisis realizados
- 6) Comprensión general de los objetivos y resultados

La valoración recogerá tanto aspectos del trabajo realizado en equipo como del individual que será evaluado mediante preguntas del profesor.