

Descripción de la tarea de evaluación

La tarea que os pido para evaluar mi parte del módulo es un **cuestionario tipo test** que tiene 15 preguntas (aunque se valora sobre 10 puntos). Leedlas detenidamente, haced las comprobaciones con python que sean necesarias y cualquier duda me la comunicáis por el foro. El cuestionario es muy fácil, y en cualquier caso tenéis **dos intentos** (por si os equivocáis a la primera y queréis modificar alguna respuesta).

Entre las 15 preguntas del cuestionario se incluyen dos preguntas, la 8 y la 9, que requieren leer datos de un fichero llamado "**DatosPractica_Scoring.xlsx**" con el que tendréis que trabajar vosotros para dar las respuestas a esas dos preguntas del cuestionario. También la pregunta 15 requiere que utilicéis datos de viviendas de Madrid "**Data_Housing_Madrid.csv**". Estas tres preguntas son las que pueden llevar más tiempo porque requieren hacer el modelo de puntuación de riesgo con los datos de la práctica, y estimar matrices de pesos espaciales!! pero valen lo mismo que las demás!!!, así que si tenéis que dejar alguna sin contestarya sabéis cual dejar para el final.

Una vez finalizado el cuestionario con las respuestas que consideréis definitivas, tenéis que enviar las respuestas para su corrección. **Hasta que no confirméis el envío para corregirlas podéis entrar y salir del cuestionario tantas veces como queráis**, grabar las respuestas que llevéis hechas para continuar más adelante y cambiar las respuestas tantas veces como queráis, sin que gastéis el primer intento.

Mientras no confirméis el envío para su corrección, el cuestionario os da la opción de "volver al último intento". Ese "último" es el intento que hayáis grabado la última vez que entrasteis al cuestionario. (Es un problema de traducción del Moodle, la plataforma que utilizamos para el curso online, y no podemos corregirlo)

No hay restricción de tiempo para contestar, la única restricción es la fecha límite de entrega que especifique el coordinador de máster, que cerrará y evaluará el cuestionario con las respuestas que hayáis mandado para evaluar. Así que pensarlo bien y si no termináis el intento en una única sesión, grabar la sesión, que siempre podréis volver a entrar al cuestionario y continuar contestando (se respetan las respuestas que hubierais contestado ya en la última vez que grabasteis).

Una vez que estáis seguros de haber rellenado completamente todo el cuestionario es cuando tenéis que enviarlo para corregir (se os pedirá confirmación para el envío). En ese momento termina vuestro primer intento (ya no podréis modificar las respuestas), entonces se os evalúa el envío. Se os dice los puntos que habéis sacado en cada respuesta sobre un máximo de 1 cada pregunta individual.

Cuando veáis la corrección del primer intento, fijaos en la puntuación obtenida en cada pregunta. Una pregunta sólo estará correctamente contestada cuando obtengáis la puntuación máxima de 1. Cualquier puntuación inferior a 1 significa que tenéis algo mal en esa pregunta (o bien os habéis equivocado en alguna respuesta, o bien falta por marcar alguna respuesta correcta).

En caso de respuestas múltiples para sacar el punto es necesario marcar todas las respuestas correctas, cada respuesta múltiple correcta tiene un valor igual al valor de la nota (+1) dividido por el número total de posibles respuestas correctas.

Todas las preguntas tienen el mismo valor: 1 punto . Aunque cada pregunta se puntúa sobre 1, **la nota final del cuestionario tiene base diez** por lo que para calcular la nota final se hace una reponderación en la que cada pregunta tiene un valor final equivalente a 10/15 puntos (porque hay 15 preguntas).

NOTA IMPORTANTE: Contestar de manera incorrecta PENALIZA restando puntos hasta (-1), de forma que si hay varias posibles respuestas erróneas la penalización (-1) se reparte entre el número de posibles respuestas incorrectas (la penalización sólo actúa cuando hay respuestas correctas, de forma que el valor final de una pregunta nunca será negativo)

Una vez calificado el primer intento podéis hacer el segundo (ojo anotad las preguntas que teníais correctas en el primer intento por si acaso no se mantienen marcadas las respuestas en el segundo intento).

La nota final del cuestionario será la mayor de las dos notas conseguidas en cada uno de los dos intentos

Una vez finalizada la fecha de entrega y cerrado el cuestionario para todo el alumnado podéis volver a entrar en vuestros ejercicios para comprobar cuál era la respuesta correcta

En algunas preguntas tendréis que introducir números **cuidado con los decimales!!!**. **Sobre todo los que utilicéis Windows y tengáis teclado español tendréis que utilizar la “coma” decimal, y NO el “punto” decimal** (para saber si lo estáis haciendo bien o no sólo tenéis que fijaros si os da error cuando intentáis pasar a la siguiente pregunta)

Información de los datos para responder a las preguntas 8 y 9 del cuestionario

Los datos para realizar el modelo de puntuación que se pide en estas preguntas del cuestionario están en el fichero Excel **DatosPráctica_Scoring.xlsx** disponible en la carpeta de archivos de clase. En este fichero Excel **DatosPráctica_Scoringxlsx** tenéis la siguiente información (también la tenéis en la hoja **Descripción datos** del fichero excel):

- Default: Variable objetivo, indica si un cliente ha hecho impago durante la ventana de observación. Default = 1 if defaulted 0 if not (observed only when Cardhldr = 1)

- Cardhldr=1: Clientes aceptados, son clientes a los que se le ha dado la tarjeta de crédito (Cardhldr=1) y sobre los que sabemos si han impagado alguna vez o no (default= 1 o 0 respectivamente).
- Cardhldr=0: clientes rechazados, son clientes que solicitaron un crédito pero a los que se negó su solicitud (Cardhldr=0), son clientes a los que no se les concedió la tarjeta, y por tanto no sabemos si hubieran impagado o no (default= na) porque no se les concedió la tarjeta de crédito.
- (Cardhldr=na): solicitud de nuevos clientes que solicitan una tarjeta de créditos y que hay que puntuar. Estos clientes están al final del archivo, son los 34 nuevos potenciales clientes (Cardhldr=na) con identificador de cliente desde 1286 hasta 1319, que están solicitando una tarjeta de crédito. Vosotros tendréis que puntuar su calidad crediticia y concederles o no el préstamo.

Para construir vuestro modelo tenéis las siguientes posibles variables explicativas o predictoras del riesgo:

- ID: identificador de cada solicitante de tarjeta
- Age = Age n years plus twelfths of a year
- Income = Yearly income (divided by 10,000)
- Exp_Inc = Ratio of monthly credit card expenditure to yearly income
- Avgexp = Average monthly credit card expenditure
- Ownrent = 1 if owns their home, 0 if rent
- Selfempl = 1 if self employed, 0 if not.
- Depndt = number of dependents (personas a cargo)
- Inc_per = Income divided by (1+number of dependents)
- Cur_add = months living at current address
- Major = number of major credit cards held
- Active = number of active credit accounts

Para la pregunta 8 se pide considerar una **muestra específica de datos** y calcular el estadístico del Valor de Información, por favor, lea bien el enunciado de esta pregunta 8.

Para la pregunta 9 se pide la construcción del modelo completo de puntuación. Con toda esta información, y considerando tanto a los clientes **Aceptados**, como a los **Rechazados** (haciendo **inferencia de rechazados**), la pregunta 9 se os pide realizar un pronóstico para las solicitudes de los nuevos clientes (los clientes con un ID desde el 1286 hasta 1319). Se recomienda realizar un análisis exploratorio previo para depurar los datos que se van a utilizar en la estimación de los modelos (**depuración** de datos atípicos, variables con poca variabilidad, transformación de variables), y dividir la muestra para entrenamiento y test para validar sus modelos).

Información de los datos para responder a la pregunta 15 del cuestionario

Para contestar a la pregunta 15 hay que trabajar con los **datos de viviendas** de la ciudad de Madrid tomados del artículo de José María Montero, Román Mínguez y Gema Fernández-Avilés (2018), Housing price prediction: parametric vs semiparametric spatial hedonic models J. Geogr Systems, vol 20, pp 27-55 (<https://doi.org/10.1007/s10109-017-0257-y>).

El objetivo es cuantificar la posible autocorrelación espacial en el precio por metro cuadrado del "centro histórico" de Madrid. Por tanto habrá que hacer una selección para usar sólo los datos del centro histórico.

Los datos están en el archivo **`Data_Housing_Madrid.csv`** disponible en la carpeta de archivos de clase. Las variables que nos van a interesar son:

- ***house.price***: precio de la vivienda (euro/m2)
- ***historical***: indica si una vivienda pertenece o no al casco histórico de Madrid
- ***longitude***: longitud en coordenadas geográficas
- ***latitude***: latitud en coordenadas geográficas

Para contestar a esta pregunta 15 habrá que seguir las instrucciones en el enunciado de la pregunta y utilizar los scripts de clase como ejemplo. Cualquier duda me la preguntáis en el foro.