**Regresión Múltiple**

**Transacciones inmobiliarias**

Cuando una familia o un individuo quiere vender su casa en muchas ocasiones lo hace a través de una agencia inmobiliaria. El agente inmobiliario ayuda a decidir el precio inicial de la casa. La familia firma un contrato, llamado acuerdo de venta con el agente. El contrato afirma que, si el agente encuentra un comprador de la casa por el precio inicial acordado, entonces la familia venderá la casa por ese precio. El precio real al que se vende la casa se llama precio de venta, y en la mayoría de los casos, menor que el precio inicial.

Una parte importante del trabajo de un agente inmobiliario es ayudar a fijar el precio inicial. El agente analizará las características de la casa, incluyendo el tamaño, localización, edad, el terreno que ocupa la casa (tamaño de la parcela), el número de dormitorios y baños, y si la casa tiene otras características deseables (sótano, garaje, chimenea, electrodomésticos, etc.). A la hora de fijar el precio, el agente considerará casas similares que se hayan vendido recientemente.

Los agentes inmobiliarios reúnen esta información en un catálogo que sería utilizado por otros agentes que ayudan a familias interesadas en comprar una casa.

**Modelos para la tasación de casas**

Las características de una casa influyen de forma muy importante en el precio de esta. La dirección del efecto de la mayoría de estas características es obvia: las casas grandes se venden más caras que las pequeñas, las casas con más dormitorios y baños tiene un precio más alto que las que tienen menos, la presencia de un garaje tiende a elevar el precio de la vivienda, etc. Pero, estas características no dan una predicción perfecta del precio de la casa. En parte, esto es debido a que algunas características no aparecen reflejadas en la información que ofrece el agente inmobiliario; por ejemplo, si la calle en la que se encuentra la casa es ruidosa o no, el estado actual de la casa, etc. Incluso en el caso de tener una lista detallada de las características no tendríamos una predicción perfecta del precio ya que éste se ve afectado por las personas que compran la casa. Si una familia ve una casa que se adapta a sus necesidades la comprará rápidamente y probablemente por el precio inicial o incluso por más, mientras que, si una casa ha estado a la venta durante mucho tiempo y la familia necesita hacer reforma para adaptarla a sus necesidades, probablemente la comprará por un precio inferior a la inicial.

1Esto es lo que ocurre en la mayoría de los casos, pero en algunos casos los individuos o familias venden directamente sus casas sin intermediarios

**Modelos hedónicos**

Un modelo hedónico es un modelo de regresión el cual relaciona el precio actual de una casa con sus características. Estos modelos tienen usos importantes:

Los agentes inmobiliarios los utilizan para decidir el precio inicial de una casa.

En la mayoría de los países, los propietarios de una vivienda pagan impuestos por ella. Estos impuestos se basan en una estimación del precio de la vivienda que se actualiza cada año. El gobierno utiliza modelos hedónicos para estimar el verlos de las viviendas. Estos modelos han de dar buenas estimaciones, ya que, si subestiman el precio, está perdiendo dinero, y si es demasiado alto, tendrá problemas con los propietarios.

**Los datos**

Utiliza el fichero Precioscasas2014.txt que nos servirá para predecir el precio mediano de viviendas en una cierta zona. Contiene 14 variables y la descripción de estas se encuentra más adelante.

**Ejercicio**

Supongamos que trabajas para una empresa consultora a la que se ha pedido que construya y estime un modelo hedónico para una ciudad. Consistirá en un modelo de regresión con el precio mediano como variable repuesta y las características de la casa como variables explicativas.

El objetivo del modelo es predecir el precio de venta de las viviendas que tienen la misma combinación de características. Se pueden construir muchos modelos. Hay que tener en cuenta que no sólo hay que elegir las variables explicativas que se han de incluir en el modelo.

Por ejemplo: la variable respuesta hay que expresarla tal y como está, en términos de logaritmos, ¿o con otra transformación?

Es necesario pensar en todos los problemas técnicos que se pueden presentar en un modelo de regresión. Para este conjunto de datos, multicolinealidad, valores extremos y varianza no constante pueden ser un problema.

Descripción de las variables

Variable descripción

-----------------------------------------------------------------------------------

CRIME tasa de crimen

ZONE proporción terreno residencial

INDUS proporción de terreno industrial

RIV variable dicotómica (=1 zona con rio; 0 sin rio)

NOX concentración de óxido nitroso

RM número medio de habitaciones

AGE proporción de casa construidas antes de 1940

DIST distancia a oficinas de empleo

RAD índice de acceso a carreteras de circunvalación

TAX tasa de impuestos sobre el precio de la casa por cada $10,000 PTRATIO ratio estudiante/maestro en las escuelas

B 1000\*(Bk-0.63)^2 donde Bk es la proporción de inmigrantes LSTAT porcentaje de población de clase baja

MEDV valor mediano de las casas en miles de dólares

----------------------------------------------------------------------------------

MEDV es la variable respuesta, las otras son variables explicativas

Analice las siguientes cuestiones:

1. Los pasos para construir el análisis (análisis exploratorio, selección de variables, ajuste del modelo, etc.).
2. Mejor modelo/modelos seleccionado
3. Bondad de ajuste del modelo (ver si se cumplen las hipótesis, R^2 ajustado, etc.).
4. Contestar a las siguientes cuestiones: ¿Cuánto aumenta/disminuye el precio de una casa si tiene el río cerca? Usando el modelo, da una predicción para el modelo de una casa con los siguientes valores de las variables explicativas (sólo utiliza aquellas que hayas incluido en tu modelo): CRIME=0.003 ZONE=18 INDUS=2.1 RIV=0 NOX=50.80 ROOM=5.575 AGE=60.2 DIST=4.19 RAD=1 TAX=293 PTRATIO=16.3 B=321.9 LSTAT=2.98.
5. ¿Es posible construir una matriz de confusión para regresiones múltiples? En caso afirmativo, hágalo para los datos de entrenamiento.