**Un conjunto de letras negras en un fondo blanco

Descripción generada automáticamente con confianza bajaUn conjunto de letras negras en un fondo blanco

Descripción generada automáticamente con confianza bajaUniversidad Autónoma de Coahuila**

**Facultad de Ingeniería Mecánica y Eléctrica**

**Modelos Computacionales**

**“Gráficos de predicción de datos”**

**Ing. Jesús Homero Carmona Mendoza**

**Daniela López Rodríguez**

**7°A ISC T.M.**

**Agosto – diciembre 2022**

**20/10/2022**

**Gráficos ICE para interpretar modelos predictivos**

Paquete ICE box

El paquete ICE box contiene funciones que permiten calcular, explorar y representar gráficos ICE para cualquier tipo de modelo predictivo.

Curvas ICE

Se entrena un modelo predictivo de tipo [Random Forest](https://rpubs.com/Joaquin_AR/255596) con el objetivo de predecir en función de todas las demás variables disponibles.

Una vez entrenado el modelo, con la función ice() se obtiene el gráfico ICE de cualquiera los predictores. Los principales argumentos de esta función son:

* object: modelo del cual se quieren obtener las curvas ICE.
* X: valor de los predictores con los que se ha entrenado el modelo.
* y: valor de la variable respuesta con la que se ha entrenado el modelo. Se emplea para identificar el rango del eje y.
* predictor: nombre o posición del predictor para el que se quiere obtener el gráfico ICE.
* frac\_to\_build: fracción de observaciones de entrenamiento que se incluyen en el gráfico ICE. Por defecto se emplean todas (frac\_to\_build = 1) pero, si el set de datos es muy grande, se recomienda reducirlo. La selección se hace de forma que se incluya aproximadamente todo el rango de valores observado en el entrenamiento.
* indices\_to\_build: índices de las observaciones que se incluyen en el gráfico ICE. Es una alternativa no aleatoria a frac\_to\_build. No pueden emplearse ambos argumentos a la vez.
* num\_grid\_pts: número de puntos dentro del rango del predictor empleado para construir la curva ICE. Por defecto, se utilizan todos los valores del predictor observados en los datos de entrenamiento del modelo.
* predictfcn: función opcional que acepta dos argumentos, un modelo (object) y un conjunto de datos (newdata), y devuelve un vector con las predicciones. Gracias a esta función se pueden obtener los gráficos ICE de cualquier modelo. Si este argumento no se especifica, se intenta encontrar automáticamente la función predict() correspondiente a la clase del modelo pasado a la función ice().

El objeto devuelto por ice() puede graficarse empleando la función plot().

Gráfico

Descripción generada automáticamente

Cada curva del grafico anterior (curva ICE) muestra el valor predicho de la variable respuesta para cada observación con forme se va aumentando el valor de Age y manteniendo constantes el resto de predictores en su valor observado. La curva resaltada en amarillo se corresponde con la curva PDP, que es la variación promedio de todas las observaciones. Además, el gráfico incluye puntos que representan el verdadero valor de Age de cada observación.

La gran mayoría de las curvas son planas, lo que apunta a que, en la mayor parte de los casos, la antigüedad de la vivienda apenas influye en su precio. Sin embargo, puede apreciarse que, unas pocas observaciones, presentan una ligera tendencia de subida o bajada.

Curvas ICE centradas

Cuando los valores observados de la variable respuesta se acumulan en una región pequeña, el solapamiento de las curvas puede hacer difícil distinguir qué observaciones que se escapan de la tendencia general. Para evitar este problema, se puede recurrir a los gráficos ICE centrados (c-ICE). Los gráficos c-ICE se obtienen igual que los gráficos ICE con la única diferencia de que, a cada una de las curvas, se les resta un valor de referencia, normalmente el valor predicho para el mínimo observado del predictor. De esta forma, se consigue que todas las curvas tengan su origen en el 0.

Gráfico

Descripción generada automáticamente

Con esta nueva representación puede observarse con más claridad que, aunque la mayoría de las observaciones se mantienen constantes, algunas tienen un claro patrón divergente (para algunas el precio incrementa con la antigüedad y en otras disminuye). Tal y como se describe más adelante, esto suele ser un indicativo de que el predictor Age interacciona con otros predictores. El eje vertical de la izquierda muestra el %% de desviación respecto al rango de yy.

Derivada de las curvas ICE

Si la relación existente entre la variable respuesta y el predictor estudiado es independiente del resto de predictores del modelo, entonces, las curvas del gráfico ICE comparten una misma forma y son paralelas las unas a las otras (la única diferencia es un desplazamiento en el eje vertical). Este comportamiento puede resultar complicado de validar visualmente cuando las curvas se superponen. Una forma de facilitar la identificación de interacciones entre predictores es representando las derivadas parciales de las curvas ICE (d-ICE). Si no existe ninguna interacción, todas las curvas son aproximadamente paralelas, sus derivadas aproximadamente iguales y, por lo tanto, el gráfico de derivadas muestra una única recta. Si existen interacciones, entonces, la representación de las derivadas parciales es heterogénea.

Gráfico, Gráfico de líneas

Descripción generada automáticamente

El gráfico sugiere que, cuando la antigüedad de la vivienda es inferior a 60 años, las derivadas parciales son ≃0≃0, por lo que no hay interacciones. Superados los 60 años, hay observaciones cuyas derivadas parciales se desvían sustancialmente de 0, indicando que, a partir de este valor, el predictor Age interacciona con otros predictores.

En la zona inferior del gráfico se muestra la desviación estándar de las derivadas parciales en cada punto, lo que facilita encontrar regiones de alta heterogeneidad (regiones de interacción).

Versión ggplot2

Toda la información presente necesaria para crear el gráfico está contenida en los objetos ice y dice, por lo que puede reproducirse la misma representación con ggplot2.

ICE de modelos H2O

La librería [H2O](http://docs.h2o.ai/h2o/latest-stable/h2o-docs/index.html) se ha convertido en un referente para modelos de machine learning. Como los modelos de H2O no se crean en el entorno de R, sino en un clúster de Java, para obtener gráficos ICE son necesarios algunas modificaciones.

* Definir una función predict que envíe los nuevos datos al clúster H2O en el formato adecuado, y que devuelva un vector con un valor numérico.
* Pasar los datos en formato h2o de nuevo a data.frame.
* Desactivar la paralelización. Por alguna razón, la función furture\_map genera un error al interactuar con H2O.

Debido a que los datos se tienen que transferir continuamente entre el clúster H2O y el entorno de R, la creación de los gráficos ICE aquí propuesta puede resultar muy lenta. Para una implementación más eficiente, conviene utilizar el paquete PDP tal como se describe en el documento [Machine Learning con H2O y R](https://rpubs.com/Joaquin_AR/406480).

Se inicia un clúster H2O y se ajusta un modelo GBM empleando la función h2o.gbm. Se puede encontrar información más detallada sobre el uso de esta libraría en el documento [Machine Learning con H2O y R](https://rpubs.com/Joaquin_AR/406480).

**Predecir tendencias de datos**

Puede extender una serie de valores que se ajusten a una tendencia lineal simple o a una tendencia de crecimiento exponencial mediante el controlador de relleno.

Use los procedimientos siguientes para obtener información sobre cómo mostrar y predecir tendencias y crear una proyección.

Crear una serie de tendencias a partir de los datos

Crear una serie de tendencias lineales que se ajusten mejor

En una serie de tendencia lineal, el valor de paso (la diferencia entre el valor inicial y el siguiente valor de la serie) se agrega al valor inicial y, a continuación, se agrega a cada valor posterior.

1. Seleccione al menos dos celdas que contengan los valores iniciales de la tendencia.

Para aumentar la precisión de la serie de tendencias, seleccione más de dos valores iniciales.

1. Arrastre el controlador de relleno en la dirección de aumentar o disminuir valores.

Por ejemplo, si seleccionó las celdas C1:E1 y los valores iniciales de estas celdas son 3, 5 y 8, arrastre el controlador de relleno a la derecha para rellenar con valores de tendencia crecientes o arrástrelo a la izquierda para rellenar con valores decrecientes.

Crear una serie de tendencias de crecimiento exponencial

En una serie de tendencia de crecimiento, el valor inicial se multiplica por el valor de paso para obtener el siguiente valor de la serie. El producto resultante y cada producto posterior se multiplican por el valor de paso.

1. Seleccione al menos dos celdas que contengan los valores iniciales de la tendencia.

Para aumentar la precisión de la serie de tendencias, seleccione más de dos valores iniciales.

1. Mantenga presionado control y arrastre el controlador de relleno en la dirección que desea rellenar con valores crecientes o decrecientes.

Por ejemplo, si seleccionó las celdas C1:E1 y los valores iniciales de estas celdas son 3, 5 y 8, arrastre el controlador de relleno a la derecha para rellenar con valores de tendencia crecientes o arrástrelo a la izquierda para rellenar con valores decrecientes.

1. Suelte CONTROL y el botón del mouse y, a continuación, en el menú contextual, haga clic en Tendencia de crecimiento.
2. Excel calcula automáticamente la tendencia de crecimiento y continúa la serie en las celdas seleccionadas.

Mostrar la serie de tendencias con una línea de tendencia en un gráfico

Puede usar línea de tendencia y media móviles gráficos bidimensionales para mostrar gráficamente tendencias y analizar problemas de predicción (también conocido como análisis de regresión). Una media móvil suaviza las fluctuaciones de los datos y muestra el patrón o tendencia con mayor claridad. Mediante el análisis de regresión, que calcula la relación entre variables, puede extender una línea de tendencia en un gráfico más allá de los datos reales para predecir valores futuros.

1. En el gráfico, haga clic en la serie de datos a la que desea agregar una línea de tendencia o una media móvil.
2. En la pestaña Diseño de gráfico, haga clic en Agregar elemento de gráfico y seleccione Línea de tendencia.
3. Haga clic en el tipo de línea de tendencia o media móvil que desea usar.
4. Haga clic en Aceptar.