**SISTEMAS INTELIGENTES**

**PROYECTO 3**

*Universidad de Deusto*

*Grado en Ingeniería Informática + Transformación Digital de la Empresa – Sistemas Inteligentes*

Profesor: Roberto Carballedo

Alexandru Nitu & Iñigo de Dios Gómez

Contenido

**No se encontraron entradas de tabla de contenido.**

Regresión Lineal

La regresión lineal es un modelo matemático usado en la Inteligencia Artificial para aproximar la relación de dependencia entre variables almacenadas en un conjunto de datos, y variables aleatorias, con el fin de predecir datos aleatorios. A continuación se muestra un ejemplo visual en modo gráfico de un modelo de Regresión Lineal:

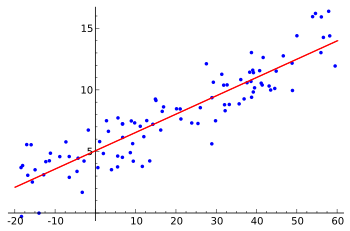


Ilustración 1: Ejemplo de una regresión lineal con una variable dependiente y una variable independiente.

Código – Análisis del mejor modelo

El código, principalmente genera 10 modelos de regresión lineal, calcula el error medio de todos ellos y selecciona el mejor modelo.

A continuación se procede a analizar los parámetros del mejor modelo.

Texto

Descripción generada automáticamente

Ilustración 2: Resultados mejor modelo

Dependencia lineal:

Analizamos las distintas variables y la dependencia lineal de las mismas. Principalmente, usamos en el eje x la edad y en el eje y el coste. Mediante la función *scatter.smooth()* obtenemos la siguiente gráfica comparativa.

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

Ilustración 3: scatter.smooth() entre edad y coste

Como podemos observar, y como bien es de esperar, ciertamente los costes tienen dependencia lineal con la edad, ya que cuanto mayor es la edad de una persona, más tendencia tiene a necesitar los servicios sanitarios.

Correlación:

Analizaremos la correlación entre cada atributo numérico y los costes, mediante la función *cor(data$y, data$x)*, para cuantificar cuánto están relacionadas las variables entre ellas.

* Edad: 0.29 Correlación normal.
* BMI: 0.19 Correlación baja.
* Hijos: 0.06 Correlación baja.

Como podemos ver, la correlación no es muy buena entre los distintos atributos, lo que quiere decir que las variables no están demasiado relacionadas.

Idoneidad de los atributos:

Como podemos observar, no todos los atributos son numéricos. Sin embargo, podríamos decir que tampoco es necesario convertirlos a modo numérico, ya que el análisis que hacemos a estos datos no dependen precisamente de estos atributos.

Significación estadística:

La significación indica si un resultado es improbable que haya sido debido al azar. En nuestro caso, los resultados de los modelos nos han dado un valor *p-value*, que es el que nos indica la significación del mismo, de 2.2e-16. El valor debe ser siempre inferior o igual a 0.05, por lo que nuestro modelo lo cumple: es improbable que los resultados de nuestro modelo sean debidos al azar.

Cuestiones

Si una persona deja de fumar, ¿cuánto se reduciría el coste?

Una persona que fuma, según se estima, tiene un coste medio de 32.555$, y una persona que no fuma, un coste medio de 8.439$. Por lo tanto, se puede decir que el coste se reduce en aproximadamente 24.115$.

¿Cuánto aumentaría el coste si una persona empieza a fumar?

La respuesta es la misma que en la anterior cuestión, pero de forma inversa. Si una persona empieza a fumar, el coste medio aumentaría en aproximadamente 24.115$.

¿Quiénes son las 3 personas cuyo coste aumentará más en 5 años?

No conseguimos una forma de encontrar la solución.