# 1 Modelo de regresión según la ecuación normal

Comparar la estimación del porcentaje de grasa corporal obtenida aplicando regresión

lineal univariable y multivariable mediante el uso de la ecuación normal normalEqn(X,y),

mostrando en los siguientes apartados el error absoluto medio alcanzado utilizando el

propio conjunto de entrenamiento como conjunto de test.

1. Mediante el conjunto completo de atributos.

­Listado de errores cometidos para regresión lineal univariable:

|  |  |
| --- | --- |
| MÉTRICA | Error cometido |
| DENSIDAD DETERMINADA POR PESAJE BAJO EL AGUA | 0.356665 |
| CIRCUNFERENCIA DEL ABDOMEN | 3.916310 |
| CIRCUNFERENCIA DEL PECHO | 4.863564 |
| CIRCUNFERENCIA DE LA CADERA | 5.250477 |
| PESO EN LIBRAS | 5.318045 |
| CIRCUNFERENCIA DEL MUSLO | 5.592949 |
| CIRCUNFERENCIA DE LA RODILLA | 5.707896 |
| CIRCUNFERENCIA DEL Biceps extendido | 5.939533 |
| CIRCUNFERENCIA DEL cuello | 5.962886 |
| CIRCUNFERENCIA DE LA MUÑECA | 6.380314 |
| CIRCUNFERENCIA DEL ANTEBRAZO | 6.422515 |
| AÑOS | 6.550792 |
| CIRCUNFERENCIA DEL TOBILLO | 6.573820 |
| ALTURA EN PULGADAS | 6.856833 |

Error cometido utilizando regresión lineal multivariable

Error absoluto medio: 0.480197

1. Listar, de mejor a peor, los resultados logrados con cada atributo individualmente.

|  |  |
| --- | --- |
| Métrica | Error cometido |
| DENSIDAD DETERMINADA POR PESAJE BAJO EL AGUA | 0.356665 |
| CIRCUNFERENCIA DEL ABDOMEN | 3.916310 |
| CIRCUNFERENCIA DEL PECHO | 4.863564 |
| CIRCUNFERENCIA DE LA CADERA | 5.250477 |
| PESO EN LIBRAS | 5.318045 |
| CIRCUNFERENCIA DEL MUSLO | 5.592949 |
| CIRCUNFERENCIA DE LA RODILLA | 5.707896 |
| CIRCUNFERENCIA DEL Biceps extendido | 5.939533 |
| CIRCUNFERENCIA DEL cuello | 5.962886 |
| CIRCUNFERENCIA DE LA MUÑECA | 6.380314 |
| CIRCUNFERENCIA DEL ANTEBRAZO | 6.422515 |
| AÑOS | 6.550792 |
| CIRCUNFERENCIA DEL TOBILLO | 6.573820 |
| ALTURA EN PULGADAS | 6.856833 |

1. Empleando, esta vez, los cinco primeros atributos del ranking elaborado en el apartado anterior.

Regresión lineal multivariable utilizando los 5 mejores atributos obtenidos en el apartado anterior.

Error absoluto medio:0.455011

1. Justificar, brevemente, los resultados obtenidos en los tres apartados anteriores.

Como observamos, la mejor variable seria la densidad determinada por el pesaje del agua. Esto es debido a que, a diferencia de las otras variables, tienen valores muy específicos y es muy sencillo para el modelo realizar los cálculos. En las otras variables vemos que para valores que son muy parecidos o iguales, la variable clase obtiene valores muy diferentes, de ahí que el error sea mayor.

Con respecto a realizar los cálculos con las 5 mejores variables, vemos que obtenemos unos resultados muy prometedores que cuando los obteníamos realizando regresión lineal univariable.

1. Repetir los apartados anteriores generando el modelo con un conjunto de entrenamiento formado por el 70% de las filas escogidas de manera aleatoria, y mostrando el error utilizando el conjunto de test formado por el 30% restante. Igualmente, compare estos nuevos resultados con los anteriores.

|  |  |
| --- | --- |
| MéTRICA | Error cometido |
| DENSIDAD DETERMINADA POR PESAJE BAJO EL AGUA | 0.407787 |
| CIRCUNFERENCIA DEL ABDOMEN | 3.885605 |
| CIRCUNFERENCIA DEL PECHO | 4.846584 |
| CIRCUNFERENCIA DE LA CADERA | 5.166225 |
| PESO EN LIBRAS | 5.563063 |
| CIRCUNFERENCIA de la rodilla | 5.694077 |
| CIRCUNFERENCIA DEl muslo | 5.780018 |
| CIRCUNFERENCIA DEL cuello | 6.115151 |
| años | 6.229067 |
| CIRCUNFERENCIA DE LA MUÑECA | 6.387144 |
| CIRCUNFERENCIA DEL biceps extendido | 6.394477 |
| CIRCUNFERENCIA DEL TOBILLO | 6.472723 |
| ALTURA EN PULGADAS | 6.736956 |
| Circunferencia del antebrazo | 6.839974 |

Como observamos, algunas posiciones del ranking han cambiado por el hecho de realizar la división de los datos y hemos obtenido un ligero aumento en el error de los datos, pero las 5 mejores posiciones del ranking se mantienen

Error cometido utilizando regresión lineal multivariable

Error absoluto medio: 0.695687

Regresión lineal multivariable utilizando los 5 mejores atributos obtenidos en el apartado anterior.

Error absoluto medio: 0.548348

# 2 Descenso del gradiente

Utilizando el mismo conjunto de entrenamiento y de test creado en el apartado e) del ejercicio anterior, obtenga nuevamente los modelos con el conjunto completo de datos y con el conjunto formado con los cinco mejores atributos, utilizando esta vez el descenso del gradiente. Realice diferentes pruebas variando los parámetros alpha y número de iteraciones hasta obtener en cada caso un error cercano al obtenido anteriormente con la ecuación normal.

Estos son los errores cometidos usando todos los atributos teniendo un conjunto de entrenamiento del 70% y de test del 30% restante:

|  |  |  |
| --- | --- | --- |
| Alpha | ITERACIONES | ERROR |
| 0.02 | 200 | 0.181339 |
| 0.01 | 500 | 0.130622 |
| 0.02 | 500 | 0.086788 |

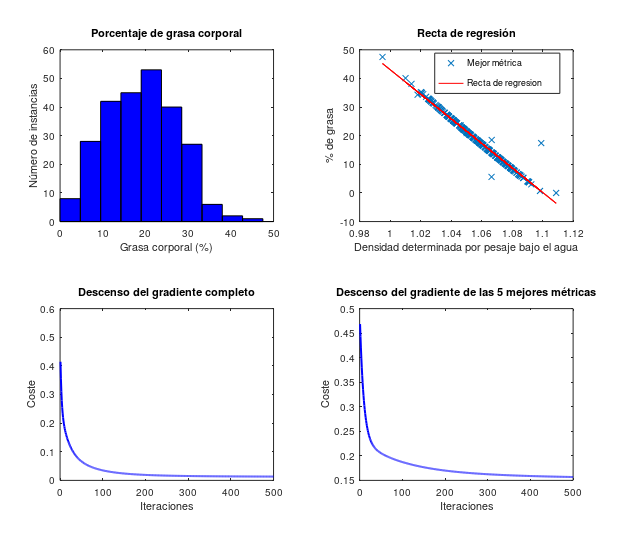
Aquí la lista de los errores cometidos en orden ascendente para Alpha a 0.02 y 500 iteraciones:

|  |  |
| --- | --- |
| Metrica | Error cometido |
| DENSIDAD DETERMINADA POR PESAJE BAJO EL AGUA | 0.031424 |
| AÑOS | 0.442317 |
| PEso en libras | 0.556667 |
| Altura en pulgadas | 0.585360 |
| CIRCUNFERENCIA DEL CUELLO | 0.605222 |
| CIRCUNFERENCIA DEL PECHO | 0.642611 |
| CIRCUNFERENCIA DEl abdomen | 0.678545 |
| CIRCUNFERENCIA DE la cadera | 0.696726 |
| CIRCUNFERENCIA DEL muslo | 0.747904 |
| CIRCUNFERENCIA de la rodilla | 0.750589 |
| CIRCUNFERENCIA del tobillo | 0.758716 |
| CIRCUNFERENCIA DEL biceps extendido | 0.781535 |
| CIRCUNFERENCIA DEl antebrazo | 0.813796 |
| CIRCUNFERENCIA DE la muñeca | 0.815482 |

En cuanto a las 5 mejores métricas, el error cometido es de 0.426162

En nuestro caso, parece ser que una vez normalizados los datos para el descenso con Alpha no desorbitadamente bajo ni muchas iteraciones, los errores cometidos han disminuido bastante.

# 3 Visualizar datos



**Gráfica 1:** Histograma del porcentaje de grasa corporal con todas las instancias del

conjunto de datos.

**Gráfica 2:** Representación de los datos del mejor atributo obtenido (densidad determinada por pesaje bajo el agua) frente al porcentaje de grasa, igualmente con recta de regresión incluida.

**Gráfica 3:** Gráfica de convergencia para el modelo obtenido en el ejercicio 2 con

el conjunto completo de atributos mediante el gradiente.

**Gráfica 4:** Gráfica de convergencia para el modelo obtenido en el ejercicio 2 con

el conjunto de cinco atributos mediante el gradiente.