



Inteligencia Artificial
Grado en Ingeniería Informática en Sistemas de Información
Curso 2020/21
HOMEWORK #1: Regresión Lineal

Como hemos visto, la regresión lineal es un campo de estudio que enfatiza la relación estadística entre dos variables continuas conocidas como variables de predicción y respuesta. Cuando hay más de una variable predictora, se convierte en regresión lineal múltiple. En esta primera práctica, comenzarás a manejar conceptos de aprendizaje automático aplicando la regresión lineal a un conjunto que muestra 252 estimaciones del porcentaje de grasa corporal determinado por 14 métricas corporales (datahw1.csv):

1. Densidad determinada por pesaje bajo el agua
2. Años
3. Peso en libras
4. Altura en pulgadas
5. Circunferencia del cuello (cm)
6. Circunferencia del pecho (cm)
7. Circunferencia del abdomen (cm)
8. Circunferencia de la cadera (cm)
9. Circunferencia del muslo (cm)
10. Circunferencia de la rodilla (cm)
11. Circunferencia del tobillo (cm)
12. Circunferencia del bíceps extendido (cm)
13. Circunferencia del antebrazo (cm)
14. Circunferencia de la muñeca (cm)
15. Porcentaje de grasa corporal

Implementar en Octave de manera clara, legible y estructurada, los siguientes ejercicios.

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:

- a) Mediante el conjunto completo de atributos
- b) Listar, de mejor a peor, los resultados logrados con cada atributo individualmente.
- c) Empleando, esta vez, los cinco primeros atributos del ranking elaborado en el apartado anterior.
- d) Justificar, brevemente, los resultados obtenidos en los tres apartados anteriores.
- e) 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.



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 α y número de iteraciones hasta obtener en cada caso un error cercano al obtenido anteriormente con la ecuación normal.

3 Visualizar datos

Mostrar una figura compuesta por cuatro gráficas:

- Histograma del porcentaje de grasa corporal con todas las instancias del conjunto de datos (no solo con el 70%).
- Representación de los datos del mejor atributo obtenido en el apartado b) del ejercicio 1, frente al porcentaje de grasa, igualmente con todas las instancias. Además, incluir la recta de regresión obtenida con dicho atributo en ese apartado.
- Mostrar la gráfica de convergencia para el modelo obtenido en el ejercicio 2 con el conjunto completo de atributos mediante el gradiente.
- Mostrar la gráfica de convergencia para el modelo obtenido en el ejercicio 2 con el conjunto de cinco atributos mediante el gradiente.