Práctica 1: Regresión lineal

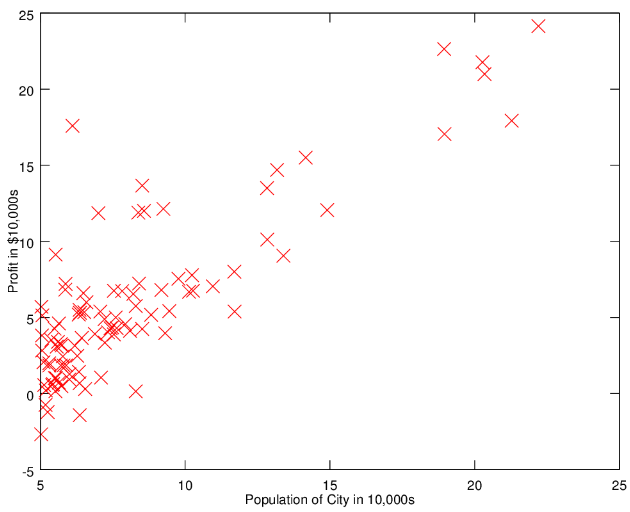
Fecha de entrega: 8 de marzo de 2018, 14.00h

Material proporcionado:

Fichero	Explicación
`ex1data1.txt`	Datos sobre los que aplicar regresión lineal con una variable.
`ex1data2.txt`	Datos sobre los que aplicar regresión lineal con varias variables.

1. Regresión lineal con una variable

En la primera parte de la práctica has de aplicar el método de regresión lineal sobre los datos del fichero `ex1data1.txt` que representan datos sobre los beneficios (segunda columna en el archivo) de una compañía de distribución de comida en distintas ciudades, en base a su población (primera columna en el archivo), como se muestra en esta figura:



Has de aplicar el método de descenso de gradiente para encontrar los parámetros θ que definen la recta que mejor se ajusta a los datos de entrenamiento. El objetivo de la regresión lineal es minimizar la función de coste:

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

donde la hipótesis $h_\theta(x)$ viene dada por el modelo lineal:

$$h_\theta(x) = \theta_0 + \theta_1 x$$

En la implementación, es una buena idea añadir un 1 como primera componente de cada ejemplo de entrenamiento x, de forma que el valor de la hipótesis $h_\theta(x)$ se pueda obtener como el producto de dos vectores:

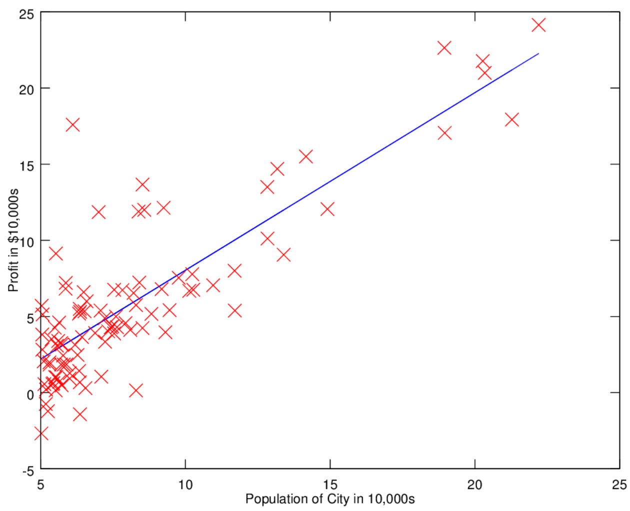
$$h_\theta(x) = \theta^T x$$

En el método de descenso de gradiente por lotes nos vamos acercando iterativamente al valor de θ que minimiza la función de coste $J(\theta)$ actualizando cada componente de θ con la expresión:

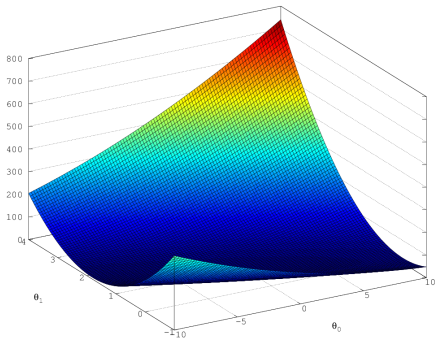
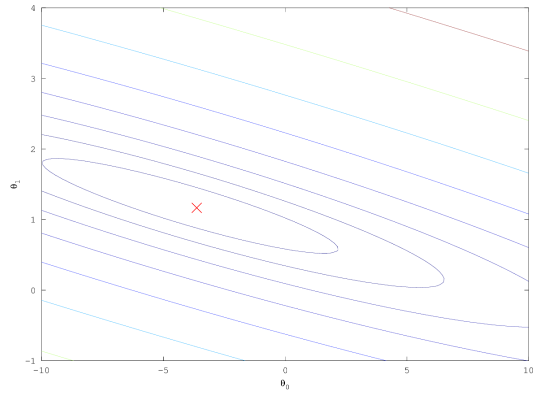
$$\theta_j := \theta_j - \alpha\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)}$$

teniendo cuidado de actualizar simultáneamente todas las componentes θ_j, es decir, computando el valor de la hipótesis $h_\theta(x^{(i)})$ en cada iteración utilizando los valores de los parámetros θ_j obtenidos en la iteración anterior.

Aplicando el método de descenso de gradiente, con unas 1500 iteraciones y un valor de $\alpha = 0{,}01$ deberías obtener unos valores para θ que definen una recta como la que se muestra en esta figura:



A modo de depuración es bueno que vayas visualizando por pantalla el valor de la función de coste $J(\theta)$ a medida que avanza el descenso de gradiente, para comprobar que efectivamente su valor decrece de manera continua. También puedes visualizar la función de coste generando gráficas como estas:

donde se muestra el resultado de llamar a las funciones *surface* y *contour* de Octave con los valores de la función de coste en el intervalo $\theta_0 \in [-10, 10]$ y $\theta_1 \in [-1, 4]$. La gráfica de contorno de la figura se ha generado utilizando una escala logarítmica para el eje z, donde se representa $J(\theta)$, con 20 intervalos entre 0.01 y 100, lo que se consigue pasando `logspace(-2, 3, 20)` como cuarto argumento de la función *contour*. En la gráfica de contorno se muestra también el mínimo obtenido por el descenso de gradiente, y se podrían mostrar otros puntos intermedios obtenidos en el proceso.

Por último, aplica el modelo obtenido a una población de 15000 y a otra de 75000.

2. Regresión con varias variables

El objetivo de la segunda parte de la práctica es aplicar el método de regresión lineal a los datos del archivo `ex1data2.m` que contienen datos sobre el precio de casas vendidas en Portland, Oregon, incluyendo para cada casa el tamaño en pies cuadrados, el número de habitaciones y el precio.

Como el rango de los distintos atributos es muy diferente (unidades en el caso del número de habitaciones y miles en el caso de la superficie) para acelerar la convergencia al aplicar el método de descenso de gradiente, es necesario que normalices los atributos, sustituyendo cada valor por el cociente entre su diferencia con la media y la desviación estándar de ese atributo en los ejemplos de entrenamiento. Para ello puedes implementar una función con la siguiente cabecera:

```
function [X_norm, mu, sigma] = normalizaAtributo(X)
```

que reciba una matriz X con los ejemplos de entrenamiento, uno por fila, y devuelva otra matriz X_norm, de las mismas dimensiones, con los valores normalizados, además de un vector mu con la media de cada atributo y otro vector $sigma$ con la desviación estándar de cada atributo. Los vectores mu y $sigma$ son necesarios para poder utilizar el modelo ajustado por el método de descenso de gradiente para hacer predicciones sobre otros ejemplos. Dados el número de habitaciones y la superficie de una nueva casa, si se quiere predecir el precio utilizando el modelo resultado del entrenamiento, es necesario normalizar esos atributos utilizando los mismos valores de media y desviación estándar con los que se normalizaron los ejemplos de entrenamiento utilizados para obtener el modelo. En Octave, la función `mean` computa la media y la función `std` la desviación estándar.

Aplica el método de descenso de gradiente a los datos normalizados para obtener el valor de los parámetros θ que minimizan la función de coste. Es conveniente realizar una implementación vectorizada de la función de coste, de la siguiente forma:

$$J(\theta) = \frac{1}{2m}(X\theta - \vec{y})^T(X\theta - \vec{y})$$

donde

$$X = \begin{bmatrix} — & (x^{(1)})^T & — \\ — & (x^{(2)})^T & — \\ & \vdots & \\ — & (x^{(m)})^T & — \end{bmatrix} \vec{y} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix}$$

A continuación, debes experimentar con el efecto de utilizar diferentes valores para la tasa de aprendizaje y construir una gráfica donde se muestre la evolución de la función de coste $J(\theta)$ a medida que avanza el descenso de gradiente, con distintos valores de tasa de aprendizaje (0.3, 0.1, 0.03, 0.01, ...).

Por último, has de resolver de nuevo el problema utilizando el método de la ecuación normal que obtiene en un sólo paso el valor óptimo para θ con la expresión:

$$\theta = (X^T X)^{-1} X^T \vec{y}$$

y recuerda que en este caso no has de normalizar los atributos.

Por último, demuestra que tus cálculos son correctos comprobando que el modelo obtenido con descenso de gradiente hace las mismas predicciones que el que resulta de la ecuación normal, aplicándolo por ejemplo a una casa con una superficie de 1.650 pies cuadrados y 3 habitaciones.

3. Entrega de la práctica

La práctica debe entregarse utilizando el mecanismo de entregas del campus virtual, no más tarde de la fecha y hora indicadas en la cabecera de la práctica.

Se entregará un único fichero en formato pdf que contenga la memoria de la práctica, incluyendo el código desarrollado y los comentarios y gráficas que se estimen más adecuados para explicar los resultados obtenidos.