

Machine Learning - Assignment 1

Manuel Allgaier

16-9-2016

note: I only included the graded questions for your convenience.

1 Question 2

2a)

Gradient Descent Algorithm

$$\begin{aligned}\theta_0 &:= \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^m (h_\theta(x^{(i)}) - y^{(i)}) \\ \theta_1 &:= \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^m (h_\theta(x^{(i)}) - y^{(i)}) x^{(i)}\end{aligned}$$

hypothesis h_θ :

$$h_\theta(x^{(i)}) = \theta_0 + \theta_1 x^{(i)}$$

learning rate $\alpha = 0.1$

number of training examples: $m = 3$

starting values: $\theta_0 = 0$ and $\theta_1 = 1$.

1st Iteration

$$\begin{aligned}\theta_0 &= 0 - 0.1 \cdot \frac{1}{3} \sum_{i=1}^3 (x^{(i)} - y^{(i)}) \\ &= -\frac{1}{30}(-3 + -2 + -4) \\ &= 0.3\end{aligned}$$

$$\begin{aligned}\theta_1 &= 1 - 0.1 \cdot \frac{1}{3} \sum_{i=1}^3 (x^{(i)} - y^{(i)})x^{(i)} \\ &= 1 - \frac{1}{30}(-9 + -10 + -24) \\ &= 1 + \frac{43}{30} = \frac{73}{30} = 2.34\end{aligned}$$

update θ_0 and θ_1 .

$$\begin{aligned}\theta_0 &:= 0.3 \\ \theta_1 &:= 2.34\end{aligned}$$

2nd Iteration

Using updated values for θ_0 and θ_1

$$\begin{aligned}\theta_0 &= 0.30 - 0.1 \cdot \frac{1}{3} \sum_{i=1}^3 (0.30 + 2.34x^{(i)} - y^{(i)}) \\ &= 0.30 - \frac{1}{30}(1.6 + 5.47 + 4.9) \\ &= 0.30 - 0.40 = -0.10\end{aligned}$$

$$\begin{aligned}\theta_1 &= 2.34 - 0.1 \cdot \frac{1}{3} \sum_{i=1}^3 (0.30 + 2.34x^{(i)} - y^{(i)})x^{(i)} \\ &= 2.34 - \frac{1}{30}(8.15 + 27.33 + 29.40) \\ &= 2.34 - 2.16 = 0.18\end{aligned}$$

update θ_0 and θ_1 .

$$\begin{aligned}\theta_0 &:= -0.10 \\ \theta_1 &:= 0.18\end{aligned}$$

Mean-Squared Error Calculate mean-squared error to evaluate hypothesis $h_\theta(x^{(i)}) = -0.10 + 0.18x^{(i)}$:

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_\theta(x^{(i)}) - y^{(i)})^2$$

insert our values $m = 3$, $\theta_0 = -0.10$ and $\theta_1 = 0.18$.

$$\begin{aligned} J(\theta_0, \theta_1) &= \frac{1}{6} \sum_{i=1}^3 (0.18x^{(i)} - 0.10 - y^{(i)})^2 \\ &= \frac{1}{6} (30.91 + 38.44 + 81.36) \\ &= 25.12 \end{aligned}$$

Result: mean-squared error= 25.12.

2b)

Convert the data to z-scores, repeat the calculations above and compare the results.

The data was converted to z-scores using <https://www.mathsisfun.com/data/standard-deviation-calculator.html> (I assume this is fine because it was not pointed out that we should do this manually).

The new values are:

x	y
-1.336	-0.981
0.267	-0.392
1.069	1.373

When repeating the calculations, I noticed that both θ values do not change, θ_0 remains 0 and θ_1 remains 1.

2 Question 4

To find a formula for θ_1 I first need to take zero as the value of the derivative relative to θ_1 of the cost function:

$$\frac{\delta}{\delta \theta_1} J(\theta_0, \theta_1) = 0$$

The following step is the same as in the gradient descent algorithm.

$$\frac{1}{m} \sum_{i=1}^m (h_\theta(x^{(i)}) - y^{(i)}) x^{(i)} = 0$$

Then simplify the equation.

$$\sum_{i=1}^m (h_\theta(x^{(i)}) - y^{(i)}) x^{(i)} = 0 \quad (1)$$

$$\sum_{i=1}^m (\theta_0 + \theta_1 x^{(i)} - y^{(i)}) x^{(i)} = 0 \quad (2)$$

$$\sum_{i=1}^m (\theta_0 x^{(i)} + \theta_1 (x^{(i)})^2 - y^{(i)} x^{(i)}) = 0 \quad (3)$$

Then break up the summation in order to get θ_1 out of it.

$$\sum_{i=1}^m (\theta_1 (x^{(i)})^2) + \sum_{i=1}^m (\theta_0 x^{(i)} - y^{(i)} x^{(i)}) = 0 \quad (4)$$

$$\theta_1 \sum_{i=1}^m ((x^{(i)})^2) + \sum_{i=1}^m (\theta_0 x^{(i)} - y^{(i)} x^{(i)}) = 0 \quad (5)$$

$$\theta_1 \sum_{i=1}^m ((x^{(i)})^2) = - \sum_{i=1}^m (\theta_0 x^{(i)} - y^{(i)} x^{(i)}) \quad (6)$$

Final equation:

$$\theta_1 = - \frac{\sum_{i=1}^m (\theta_0 x^{(i)} - y^{(i)} x^{(i)})}{\sum_{i=1}^m ((x^{(i)})^2)}$$