Assignment 1

Wendy Nieuwkamer

16-9-2016

1 Question 1

Suppose that we have historical data of result of soccer matches of teams playing against Ajax. We wantto use this information to learn to predict at a certain moment whether a team will win, lose or draw against Ajax. Our approach will be based on Machine Learning.

1.1 Define the given and the goal of the prediction task. Classify the learning task.

Given are the results of soccer matches, the goal is to decide whether a team will win, lose or draw when playing. The learning task is supervised, as we are given inputs with their wanted result. It is a classification problem as every input should be classified as one of three possibilities.

1.2 What would be the form of training data for the learning task? Give a small training set.

An example of a training set could be the following table:

team	results
1	win
2	draw
3	lose
1	lose
3	draw
4	win

Table 1: training examples

2 Question 2

Use the training data in the table to answer all the subquestions.

x	y
3	6
5	7
6	10

2.1 Manually calculate two iterations of the gradient descent algorithm for univariate linear regression function.

Initialize the parameters such that the regression function passes through the origin (0, 0) and has an angle of 45 degrees. Use a learning rate of 0.1. Give the intermediate results of your calculations and also compute the mean-squared error of the function after 2 iterations.

Gradient Descent Algorithm

Repeat until convergence {

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

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

}

The hypothesis h_θ is:

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

For this exercise the learning rate $\alpha = 0.1$ and the amount of learning examples $m = 3$ are given. The line that passes through the origin and has an angle of 45 degrees relative to the x-axis is $f(x) = x$, thus our starting values are $\theta_0 = 0$ and $\theta_1 = 1$.

First Iteration First, calculate the new values for θ_0 and θ_1.

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

$$\begin{aligned}\theta_1 &= 1 - 0.1 \cdot \frac{1}{3}\sum_{i=1}^{3}(x^{(i)} - y^{(i)})x^{(i)} \\ &= 1 - \frac{1}{30}(-9 + -10 + -24) \\ &= 1 + \frac{43}{30} = \frac{73}{30} = 2.43\end{aligned}$$

Then, update θ_0 and θ_1.

$$\begin{aligned}\theta_0 &:= 0.3 \\ \theta_1 &:= 2.43\end{aligned}$$

Second Iteration The values for m and α are still the same, but now we use the new values for θ_0 and θ_1. Again, first the calculations.

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

$$\begin{aligned}\theta_1 &= 2.43 - 0.1 \cdot \frac{1}{3}\sum_{i=1}^{3}(0.30 + 2.34x^{(i)} - y^{(i)})x^{(i)} \\ &= 2.43 - \frac{1}{30}(4.80 + 27.33 + 29.40) \\ &= 2.43 - 2.05 = 0.38\end{aligned}$$

Then, update θ_0 and θ_1.

$$\theta_0 := -0.10$$
$$\theta_1 := 0.38$$

Mean-Squared Error Now we can calculate the mean-squared error to see how good our hypothesis $h_\theta(x^{(i)}) = -0.10 + 0.38x^{(i)}$ is. The formula we use is:

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

This function is also called the cost function. We can now insert our values $m = 3$, $\theta_0 = -0.10$ and $\theta_1 = 0.38$.

$$\begin{aligned} J(\theta_0, \theta_1) &= \frac{1}{6} \sum_{i=1}^{3} (0.38x^{(i)} - 0.10 - y^{(i)})^2 \\ &= \frac{1}{6}(24.60 + 27.04 + 61.15) \\ &= 18.80 \end{aligned}$$

Thus, the mean-squared error for $m = 3$, $\theta_0 = -0.10$, $\theta_1 = 0.38$ and the given training examples is 18.80.

2.2 Convert the data to z-scores (with mean = 0, sd = 1), repeat the calculations above and compare the results.

At first I interpreted this question as: "To calculate the z-scores use mean $\mu = 0$ and standard deviation $\sigma = 1$." *Which leads to the following answer:* "The formula for calculating a z-score is:

$$Z = \frac{X - \mu}{\sigma}$$

Where X is the input value of the training example. Inserting our μ and σ into this formula gives:

$$Z = \frac{X - 0}{1} = X$$

Redoing the calculations would not be very useful as with this μ and σ the Z-score would be the same as the original input. It would be more useful to choose $\mu = 4.5$ and $\sigma = 1.125$ for x as this actually changes the values."

Then I realized that the Z-scores were supposed to have mean $\mu = 0$ and standard deviation $\sigma = 1$, which leads to the following answer: First, we have to calculate the mean μ of our training examples. The formula for this is:

$$\mu = \frac{\sum\limits_{i=1}^{m} x^{(i)}}{m}$$

Where m is the amount of training examples. So, for our x values this gives $\mu_x = 4\frac{2}{3} = 4.67$ and for the y values it gives $\mu_y = 7.67$.

Now I ran into some questions. One of them is when you apply this to your x values, do you have to do the same with your y values. And why? Also, when we calculate the standard deviation, do we calculate the Population standard deviation or the Sample standard deviation?

I decided to take the Population Standard deviation as calculated by the website `https://www.mathsisfun.com/data/standard-deviation-calculator.html`. The reason for this is that we were asked to use σ, which is only used for the population standard deviation. This gave me the value $\sigma_x = 1.247219129$ and $\sigma_y = 1.699673171$. Now applying the formula above to my values gives me the following:

x	y
-1.34	-1.00
0.27	-0.39
1.07	1.37

With these new values we repeat the same motions as in the last subsection. Again, our starting values are $\alpha = 0.1$, $\mu = 3$, $\theta_0 = 0$ and $\theta_1 = 1$.

First Iteration First, calculate the new values for θ_0 and θ_1.

$$
\begin{aligned}
\theta_0 &= 0 - 0.1 \cdot \frac{1}{3} \sum_{i=1}^{3} (x^{(i)} - y^{(i)}) \\
&= -\frac{1}{30}(-0.34 + 0.66 - 0.30) \\
&= -6.67e^{-4}
\end{aligned}
$$

$$
\begin{aligned}
\theta_1 &= 1 - 0.1 \cdot \frac{1}{3} \sum_{i=1}^{3} (x^{(i)} - y^{(i)}) x^{(i)} \\
&= 1 - \frac{1}{30}(0.46 + 0.18 - 0.32) \\
&= 1 + 0.01 = 0.99
\end{aligned}
$$

Then, update θ_0 and θ_1.

$$
\begin{aligned}
\theta_0 &:= -6.67e^{-4} \\
\theta_1 &:= 0.99
\end{aligned}
$$

I don't have time to do the second iteration and calculate the mean-squared error for the results as I did not understand the question in time.

3 Question 3

4 Question 4

Derive an equation that can be used to find the optimal value of the parameter θ_1 for univariate linear regression without doing gradient descent. This can be done by setting the value of the derivative equal to 0. You may assume that the value of θ_0 is fixed.

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

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

We have seen this derivative before as a part of the gradient descent algorithm, so the next step is easy.

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

Now, in order to simplify the equation, we first multiply both sides by m (1), write out the hypothesis h_θ (2) and write out the multiplication with $x^{(i)}$ (3).

$$\begin{aligned}
\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x^{(i)}) &= 0 &(1)\\
\sum_{i=1}^{m}((\theta_0 + \theta_1 x^{(i)} - y^{(i)})x^{(i)}) &= 0 &(2)\\
\sum_{i=1}^{m}(\theta_0 x^{(i)} + \theta_1 (x^{(i)})^2 - y^{(i)}x^{(i)}) &= 0 &(3)
\end{aligned}$$

Now, we can breakup the summation (4), factor out θ_1 (5) and the summation without θ_1 to the right side of the equation (6).

$$\begin{aligned}
\sum_{i=1}^{m}(\theta_1(x^{(i)})^2) + \sum_{i=1}^{m}(\theta_0 x^{(i)} - y^{(i)}x^{(i)}) = 0 &&(4)\\
\theta_1\sum_{i=1}^{m}((x^{(i)})^2) + \sum_{i=1}^{m}(\theta_0 x^{(i)} - y^{(i)}x^{(i)}) = 0 &&(5)\\
\theta_1\sum_{i=1}^{m}((x^{(i)})^2) = -\sum_{i=1}^{m}(\theta_0 x^{(i)} - y^{(i)}x^{(i)}) &&(6)
\end{aligned}$$

Finally, divide both sides by the left sum to get an equation for θ_1.

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