Machine Learning: Assignment 1

Leah Dickhoff
S11155515
due September 16, 2016

1.

a) Prediction task:

The given: which team plays against Ajax in the upcoming game(s), where it is played, etc.
The goal: to predict Ajax will win, lose or draw against a specific team at a specific moment

Learning task:

The given: the set of historical data of results of the soccer matches (if win/loss/draw, stadium played in, previous matches, etc.)
The goal: connect input (all the data, the 'given') and output by establishing an appropriate algorithm
Classified as: supervised learning
Supervised as: classification problem

b) One possible form of the training data could be:

opponent team	stadium (1=home, 0=away)	number of previous draws	wins	losses
FC Barcelona	0	2	0	5

2.

a) Parameters Initialization:

function $h_\theta(x^{(i)})=\theta_0+\theta_1 x^{(i)}$ passes through origin => $\theta_0 = 0$
function has an angle of 45 degrees => $\theta_1 = 1$
learning rate: $\alpha = 0.1$
number of training samples: m=3

gradient descent:

first iteration:

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

$h_\theta(3) = 3$

$h_\theta(5) = 5$

$h_\theta(6) = 6$

$$\theta_0 = \theta_0 - \alpha \times 1/m \times \sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})$$
$$= - 0.1 \times 1/3 \times ((h_\theta(3)-6) + (h_\theta(5)-7) + (h_\theta(6)-10))$$
$$= - 0.1 \times 1/3 \times (-9)$$
$$= 0.3$$

$$\theta_1 = \theta_1 - \alpha \times 1/m \times \sum_{i=1}^{m}[(h_\theta(x^{(i)}) - y^{(i)}) \times x^{(i)}]$$
$$= 1 - 0.1 \times 1/3 \times [(h_\theta(3)-6)\times 3 + (h_\theta(5)-7)\times 5 + (h_\theta(6)-10)x6]$$

$= 1 - 0.1 \times 1/3 \times (-9-10-24)$
$= 1 - 0.1 \times 1/3 \times (-43)$
$= 2.4\overline{3}$

second iteration:

$h_\theta(x^{(i)}) = \theta_0+\theta_1 x^{(i)} = 0.3 + 2.4\overline{3}\, x^{(i)}$ and

$h_\theta(3) = 7.6$

$h_\theta(5) = 12.47$

$h_\theta(6) = 14.9$

$\theta_0 = \theta_0 - \alpha \times 1/m \times \sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})$
$= 0.3 - 0.1 \times 1/3 \times [(h_\theta(3)-6) + (h_\theta(5)-7) + (h_\theta(6)-10)]$
$= 0.3 - 0.1 \times 1/3 \times (1.6 + 5.47 + 4.9)$
$= 0.3 - 0.1 \times 1/3 \times 11.97$
$= -0.099$

$\theta_1 = \theta_1 - \alpha \times 1/m \times \sum_{i=1}^{m}[(h_\theta(x^{(i)}) - y^{(i)}) \times x^{(i)}]$
$= 2.4\overline{3} - 0.1 \times 1/3 \times [(h_\theta(3)-6)\times 3 + (h_\theta(5)-7)\times 5 + (h_\theta(6)-10)x6]$
$= 2.4\overline{3} - 0.1 \times 1/3 \times (4.8 + 17.35 + 29.4)$
$= 2.4\overline{3} - 0.1 \times 1/3 \times 51.55$
$= 0.715$

concluded function:

$h_\theta(x^{(i)}) = \theta_0+\theta_1 x^{(i)} = -0.099 + 0.715 \times x^{(i)}$

Mean-squared error of the function:

$J(\theta) = 1/2m \times \sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$
$= 1/6 \times \sum_{i=1}^{m}(-0.099 + 0.715 \times x^{(i)} - y^{(i)})^2$
$= 1/6\, [(-0.099+0.715\times 3-6) + (-0.099+0.715\times 5-7) + (-0.099+0.715\times 6-10)]^2$
$= 29.4241$

b) let x'' represent the converted data, then:

mean value: $\mu = \frac{3+5+6}{3} = 4.\overline{6}$ (= mu)

$x' = x - \mu$

standard deviation: $\sigma = \sqrt{\frac{\sum(x-\mathrm{mu})^2}{number\ of\ values}} = \sqrt{\frac{(3-4.\overline{6})^2+(5-4.\overline{6})^2+(6-4.\overline{6})^2}{3}} = \frac{\sqrt{14}}{3} \approx 1.2472$

$x'' = x'/\sigma$

x	x'	x''	y
3	$-1.\overline{6}$	-1.3363	6
5	$0.\overline{3}$	0.2673	7
6	$1.\overline{3}$	1.0690	10

gradient descent:

first iteration:

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

$h_\theta(-1.3363) = -1.3363$

$h_\theta(0.2673) = 0.2673$

$h_\theta(1.0690) = 1.0690$

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

$= - 0.1 \times 1/3 \times [(-1.3363-6) + (0.2673-7) + (1.0690-10)]$

$= 0.7\overline{6}$

$\theta_1 = \theta_1 - \alpha \times 1/m \times \sum_{i=1}^{m}[(h_\theta(x^{(i)}) - y^{(i)}) \times x^{(i)}]$

$= 1 - 0.1 \times 1/3 \times [(-1.3363-6)\times 3 + (0.2673-7)\times 5 + (1.0690-10)x6]$

$= 4.6419$

second iteration:

$h_\theta(x^{(i)}) = \theta_0+\theta_1 x^{(i)} = 0.7\overline{6} + 4.6419\, x^{(i)}$ and

$h_\theta(-1.3363) = -5.4364$

$h_\theta(0.2673) = 2.0075$

$h_\theta(1.0690) = 5.7289$

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

$= 0.7\overline{6} - 0.1 \times 1/3 \times [(h_\theta(-1.3363)-6) + (h_\theta(0.2673)-7) + (h_\theta(1.0690)-10)]$

$= 0.7\overline{6} - 0.1 \times 1/3 \times [(-5.4364-6) + (2.0075-7) + (5.7289-10)]$

$= 1.45\overline{6}$

$\theta_1 = \theta_1 - \alpha \times 1/m \times \sum_{i=1}^{m}[(h_\theta(x^{(i)}) - y^{(i)}) \times x^{(i)}]$

$= 4.6419 - 0.1 \times 1/3 \times [(-5.4364-6)\times 3 + (2.0075-7)\times 5 + (5.7289-10)x6]$

$= 7.76595$

concluded function:

$h_\theta(x^{(i)}) = \theta_0+\theta_1 x^{(i)} = 1.45\overline{6}+ 7.76595 \times x^{(i)}$

Mean-squared error of the function:

$J(\theta) = 1/2m \times \sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$

$= 1/6 \times \sum_{i=1}^{m}(1.45\overline{6}+ 7.76595 \times x^{(i)} - y^{(i)})^2$

$= 1/6\ [(1.45\overline{6} + 7.76595 \times -1.3363 - 6)$

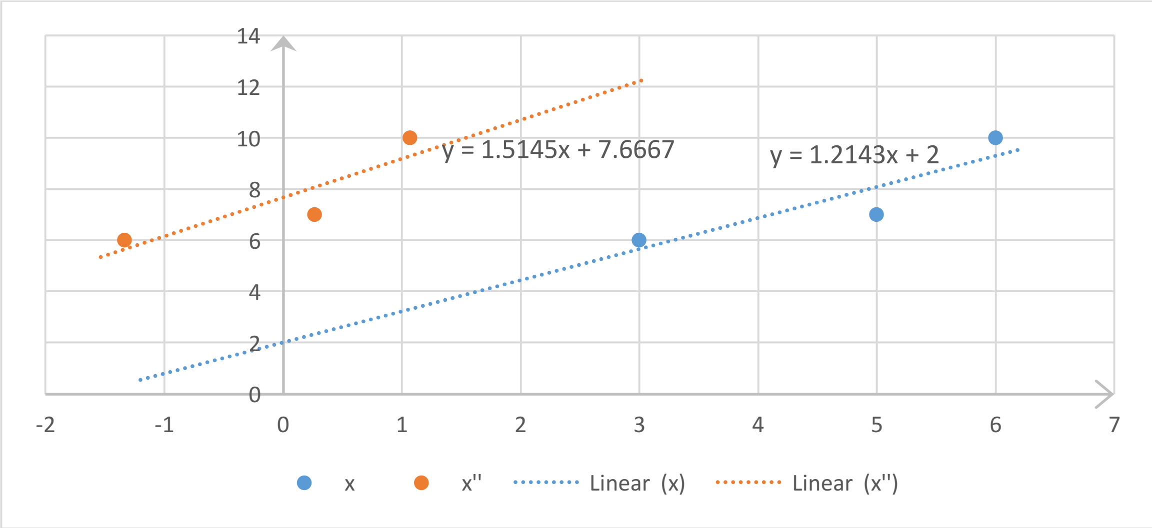
$+ (1.45\overline{6} + 7.76595 \times 0.2673 - 7)$

$+ (1.45\overline{6} + 7.76595 \times 1.0690 - 10)]^2$

$= 56.3185$

Comparison:

Let's plot the two set of points (x and x'') on the x-axis and the y values on the y-axis.



It can be concluded that the equation found using the the converted to z-scores data is more accurate (comparing it to the equation found using EXCEL), then the first set of data is.
However, the mean-squared error (MSE) of the function using the converted data is bigger than the first MSE; which can be concluded from the graph, too.

3. X1 stays fixed throughout the process.
   We use X1 and X2 in relation with the MSE to get Y.
   a) if X2 = a + b × X1
      then the MSE will decrease linearly
   b) if $X2 = a + b \times X1^2$
      then the MSE will decrease quadratically

4. Setting the derivative of $J(\theta)$ over θ_1 equal to zero:

$$\frac{\delta}{\delta\theta 1}J(\theta) = 0$$
$$\Leftrightarrow 1/2m \times \frac{\delta}{\delta\theta 1}\left[\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2\right] = 0$$
$$(\Leftrightarrow 1/2m \times \frac{\delta}{\delta\theta 1}[(h_\theta(1)-y^{(1)})^2+(h_\theta(2)-y^{(2)})^2+\ldots+(h_\theta(m)-y^{(m)})^2] = 0)$$
$$\Leftrightarrow 1/m \times \sum_{i=1}^{m}[(h_\theta(x^{(i)}) - y^{(i)}) \times x^{(i)}] = 0$$
$$\Leftrightarrow \sum_{i=1}^{m}[(\theta_0 + \theta_1 x^{(i)} - y^{(i)}) \times x^{(i)}] = 0$$
$$\Leftrightarrow \sum_{i=1}^{m}(\theta_0\, x^{(i)} + \theta_1 x^{2(i)} - y^{(i)}\, x^{(i)}) = 0$$
$$\Leftrightarrow \sum_{i=1}^{m}(\theta_0\, x^{(i)}) + \sum_{i=1}^{m}(\theta_1 x^{2(i)}) - \sum_{i=1}^{m}(y^{(i)}\, x^{(i)}) = 0$$
$$\Leftrightarrow \sum_{i=1}^{m}(\theta_1 x^{2(i)}) = \sum_{i=1}^{m}(y^{(i)}\, x^{(i)}) - \sum_{i=1}^{m}(\theta_0\, x^{(i)})$$
$$\Leftrightarrow \sum_{i=1}^{m}(\theta_1 x^{(i)}) = \sum_{i=1}^{m}(y^{(i)}) - \sum_{i=1}^{m}(\theta_0)$$
$$\Leftrightarrow m\theta_1 = \sum_{i=1}^{m}[(y^{(i)}- \theta_0)/x^{(i)}]$$
$$\Leftrightarrow \theta_1 = 1/m \times \sum_{i=1}^{m}[(y^{(i)}- \theta_0)/x^{(i)}]$$

which is an equation that gives the optimal value of the parameter θ_1 for univariate linear regression.