Leren Written assignment 2 2016-2017

Micha de Groot and Danny Dijkzeul

November 2016

1 Multiple linear regression by hand

We do two iterations of linear regression with the following data:

$$X = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 4 & 5 \\ 1 & 4 & 3 \end{bmatrix}$$

$$Y = \begin{bmatrix} 6 \\ 6 \\ 10 \end{bmatrix}$$

Our learning rate is: $\alpha = 0,01$

The initial parameters are: $\theta = \begin{bmatrix} 0.2 \\ 0.2 \\ 0.2 \end{bmatrix}$

The starting cost is: $J(\theta) = \frac{1}{6} \sum_{i=0}^{3} (\theta^T X^{(i)} - Y^{(i)})^2$

$$J(\theta) = \frac{1}{6} \sum_{i=0}^{3} (\begin{bmatrix} 0.2 & 0.2 & 0.2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - 6)^2 + (\begin{bmatrix} 0.2 & 0.2 & 0.2 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \\ 5 \end{bmatrix} - 6)^2 +$$
$$(\begin{bmatrix} 0.2 & 0.2 & 0.2 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \\ 3 \end{bmatrix} - 10)^2 = 18.27$$

Update rule:

$$\theta_j = \theta_j - \alpha \frac{1}{3} \sum_{i=0}^{3} (\theta^T X^{(i)} - Y^{(i)}) * X_j^{(i)}$$

$$\theta_0 = 0.2 - 0.01 \frac{1}{3} \sum_{i=0}^{3} (\begin{bmatrix} 0.2 & 0.2 & 0.2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - 6) * 1 +$$
$$(\begin{bmatrix} 0.2 & 0.2 & 0.2 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \\ 5 \end{bmatrix} - 6) * 1 + (\begin{bmatrix} 0.2 & 0.2 & 0.2 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \\ 3 \end{bmatrix} - 10) * 1 = 0.257$$

$\theta_1 = 0.2 - 0.01\frac{1}{3}\sum_{i=0}^{3}(\begin{bmatrix}0.2 & 0.2 & 0.2\end{bmatrix}\begin{bmatrix}1\\2\\3\end{bmatrix} - 6) * 2 +$

$(\begin{bmatrix}0.2 & 0.2 & 0.2\end{bmatrix}\begin{bmatrix}1\\4\\5\end{bmatrix} - 6) * 4 + (\begin{bmatrix}0.2 & 0.2 & 0.2\end{bmatrix}\begin{bmatrix}1\\4\\3\end{bmatrix} - 10) * 4 = 0.397$

$\theta_2 = 0.2 - 0.01\frac{1}{3}\sum_{i=0}^{3}(\begin{bmatrix}0.2 & 0.2 & 0.2\end{bmatrix}\begin{bmatrix}1\\2\\3\end{bmatrix} - 6) * 3 +$

$(\begin{bmatrix}0.2 & 0.2 & 0.2\end{bmatrix}\begin{bmatrix}1\\4\\5\end{bmatrix} - 6) * 5 + (\begin{bmatrix}0.2 & 0.2 & 0.2\end{bmatrix}\begin{bmatrix}1\\4\\3\end{bmatrix} - 10) * 3 = 0.399$

Applying for the whole θ vector gives the new θ vector:

$\theta = \begin{bmatrix}0.257\\0.397\\0.399\end{bmatrix}$

$J(\theta) = \frac{1}{6}\sum_{i=0}^{3}\begin{bmatrix}0.257 & 0.397 & 0.399\end{bmatrix}\begin{bmatrix}1\\2\\3\end{bmatrix} - 6)^2 + (\begin{bmatrix}0.257 & 0.397 & 0.399\end{bmatrix}\begin{bmatrix}1\\4\\5\end{bmatrix} -$

$6)^2 + (\begin{bmatrix}0.257 & 0.397 & 0.399\end{bmatrix}\begin{bmatrix}1\\4\\3\end{bmatrix} - 10)^2 = 11.19$

The new cost is: $J(\theta) = 11.19$

Calculating the next iteration:

$\theta_0 = 0.257 - 0.01\frac{1}{3}\sum_{i=0}^{3}(\begin{bmatrix}0.257 & 0.397 & 0.399\end{bmatrix}\begin{bmatrix}1\\2\\3\end{bmatrix} - 6) * 1 +$

$(\begin{bmatrix}0.257 & 0.397 & 0.399\end{bmatrix}\begin{bmatrix}1\\4\\5\end{bmatrix} - 6) * 1 + (\begin{bmatrix}0.257 & 0.397 & 0.399\end{bmatrix}\begin{bmatrix}1\\4\\3\end{bmatrix} - 10) * 1 = 0.30$

$\theta_1 = 0.397 - 0.01\frac{1}{3}\sum_{i=0}^{3}(\begin{bmatrix}0.257 & 0.397 & 0.399\end{bmatrix}\begin{bmatrix}1\\2\\3\end{bmatrix} - 6) * 2 +$

$(\begin{bmatrix}0.257 & 0.397 & 0.399\end{bmatrix}\begin{bmatrix}1\\4\\5\end{bmatrix} - 6) * 4 + (\begin{bmatrix}0.257 & 0.397 & 0.399\end{bmatrix}\begin{bmatrix}1\\4\\3\end{bmatrix} - 10) * 4 = 0.54$

$\theta_2 = 0.399 - 0.01\frac{1}{3}\sum_{i=0}^{3}(\begin{bmatrix}0.257 & 0.397 & 0.399\end{bmatrix}\begin{bmatrix}1\\2\\3\end{bmatrix} - 6) * 3 +$

$$(\begin{bmatrix}0.257 & 0.397 & 0.399\end{bmatrix}\begin{bmatrix}1\\4\\5\end{bmatrix}-6)*5+(\begin{bmatrix}0.257 & 0.397 & 0.399\end{bmatrix}\begin{bmatrix}1\\4\\3\end{bmatrix}-10)*3 = 0.54$$

$$\theta = \begin{bmatrix}0.30\\0.54\\0.54\end{bmatrix}$$

The final cost after two iterations is:

$$J(\theta) = \frac{1}{6}\sum_{i=0}^{3}\begin{bmatrix}0.30 & 0.54 & 0.54\end{bmatrix}\begin{bmatrix}1\\2\\3\end{bmatrix}-6)^2+(\begin{bmatrix}0.30 & 0.54 & 0.54\end{bmatrix}\begin{bmatrix}1\\4\\5\end{bmatrix}-6)^2+$$

$$(\begin{bmatrix}0.30 & 0.54 & 0.549\end{bmatrix}\begin{bmatrix}1\\4\\3\end{bmatrix}-10)^2 = 7.40$$

$$J(\theta) = 7.40$$

2 Vectorization

(a)

The number of training examples is m. The number of features is n. The dimensions of X are $m*(n+1)$.
The dimensions of Y are $m*1$.
The dimensions of θ are $(n+1)*1$

(b)

For a single feature:

$$\theta_j = \theta - \alpha\frac{1}{m}\sum_{i=0}^{m}(\theta_j * X^{(i)} - Y^{(i)}) * X_j^{(i)}$$

For all features:

$$\Theta = \Theta - \alpha\frac{1}{m}\sum_{i=0}^{m}(\Theta^T * X^{(i)} - Y^{(i)})^T * X^{(i)}$$

(c)

The hypothesis is:

$$h_\theta(X) = \Theta^T X$$

(d)

Update rule vectorized:

$$\Theta = \Theta - \alpha * \frac{1}{M}((\Theta^T X)^T - Y)^T X$$

(e)

Vectorized update rule applied to dataset:
$\alpha = 0.01$

$$X = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 4 & 5 \\ 1 & 4 & 3 \end{bmatrix}$$

$$Y = \begin{bmatrix} 6 \\ 6 \\ 10 \end{bmatrix}$$

$$\theta = \begin{bmatrix} 0.2 \\ 0.2 \\ 0.2 \end{bmatrix}$$

$$\Theta = \begin{bmatrix} 0.2 \\ 0.2 \\ 0.2 \end{bmatrix} - 0.01 * \frac{1}{3}((\begin{bmatrix} 0.2 & 0.2 & 0.2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 1 & 4 & 5 \\ 1 & 4 & 3 \end{bmatrix}) - \begin{bmatrix} 6 \\ 6 \\ 10 \end{bmatrix})^T \cdot \begin{bmatrix} 1 & 2 & 3 \\ 1 & 4 & 5 \\ 1 & 4 & 3 \end{bmatrix}$$

$$\Theta = \begin{bmatrix} 0.2 \\ 0.2 \\ 0.2 \end{bmatrix} - 0.01 * \frac{1}{3}(\begin{bmatrix} 0.6 \\ 2 \\ 2.2 \end{bmatrix} - \begin{bmatrix} 6 \\ 6 \\ 10 \end{bmatrix})^T \cdot \begin{bmatrix} 1 & 2 & 3 \\ 1 & 4 & 5 \\ 1 & 4 & 3 \end{bmatrix}$$

$$\Theta = \begin{bmatrix} 0.2 \\ 0.2 \\ 0.2 \end{bmatrix} - 0.01 * \frac{1}{3} \begin{bmatrix} -5.4 & -4 & -7.8 \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 & 3 \\ 1 & 4 & 5 \\ 1 & 4 & 3 \end{bmatrix}$$

$$\Theta = \begin{bmatrix} 0.2 \\ 0.2 \\ 0.2 \end{bmatrix} - 0.01 * \frac{1}{3} \begin{bmatrix} -17.2 \\ -58 \\ -59.6 \end{bmatrix}$$

$$\Theta = \begin{bmatrix} 0.2 \\ 0.2 \\ 0.2 \end{bmatrix} - \begin{bmatrix} -0.0574 \\ -0.1933 \\ -0.1987 \end{bmatrix}$$

$$\Theta = \begin{bmatrix} -0.2574 \\ -0.3933 \\ -0.3987 \end{bmatrix}$$

3 Logistic regression by hand

We do one iteration of logistic regression:

$$X = \begin{bmatrix} 1 & 5 & 3 \\ 1 & 5 & 3 \\ 1 & 3 & 3 \\ 1 & 2 & 4 \end{bmatrix}$$
$$Y = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$
$$\Theta = \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \end{bmatrix}$$
$\alpha = 0.01$

The hypothesis is:

$h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$

The cost function is:

$J(\theta) = -\frac{1}{m}[\sum_{i=0}^{m} y^{(i)} log(h_\theta(x^{(i)})) + (1-y) log(1-h_\theta(x^{(i)}))]$

The update rule is:

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

The initial cost is:

$J(\theta) = -\frac{1}{4}($
$log(1 - h_\theta([1,5,3]))$
$+ log(1 - h_\theta([1,5,3]))$
$+ log(h_\theta([1,3,3]))$
$+ log(h_\theta([1,2,4]))$
$)$

$= -\frac{1}{4}($
$log(1 - \frac{1}{1+e^{-4.5}})$
$+ log(1 - \frac{1}{1+e^{-4.5}})$
$+ log(\frac{1}{1+e^{-3.5}})$
$+ log(\frac{1}{1+e^{-3.5}})$
$)$

$= -\frac{1}{4}($
-4.51
-4.51
-0.03
-0.03

)
$= 2.27$

Now we do one iteration of logistic regression:

$$\Theta = \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \end{bmatrix} - 0.01\tfrac{1}{4}\sum_{i=0}^{m}(h_\Theta(x^{(i)}) - y^{(i)})x^{(i)}$$

The part that has not been filled in yet can be done more efficiently if we do not calculate this straightforward. We do it in several steps.

First we calculate all the values that have to be inserted into the sigmoid function.

$$(\Theta^T X^T)^T = \begin{bmatrix} 0.5 & 0.5 & 0.5 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 5 & 5 & 3 & 2 \\ 3 & 3 & 4 & 4 \end{bmatrix} = \begin{bmatrix} 4.5 \\ 4.5 \\ 4 \\ 3.5 \end{bmatrix}$$

Then we calculate the sigmoid function for the whole vector.

$$\begin{bmatrix} \frac{1}{1+e^{-4.5}} \\ \frac{1}{1+e^{-4.5}} \\ \frac{1}{1+e^{-4}} \\ \frac{1}{1+e^{-3.5}} \end{bmatrix} = \begin{bmatrix} 0.99 \\ 0.99 \\ 0.98 \\ 0.97 \end{bmatrix}$$

Then we subtract the whole Y vector at once

$$h_\Theta(X) - Y = \begin{bmatrix} 0.99 \\ 0.99 \\ 0.98 \\ 0.97 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.99 \\ 0.99 \\ -0.02 \\ -0.03 \end{bmatrix}$$

Plugging this back into the formula gives:

$$\Theta = \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \end{bmatrix} - 0.01\tfrac{1}{4}\sum_{i=0}^{m} \begin{bmatrix} 0.99 \\ 0.99 \\ -0.02 \\ -0.03 \end{bmatrix} x^{(i)}$$

$$\Theta = \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \end{bmatrix} - 0.0025\sum_{i=0}^{m} \begin{bmatrix} 0.99 \\ 0.99 \\ -0.02 \\ -0.03 \end{bmatrix} x^{(i)}$$

This can be simplified by multiplying with the whole X matrix if we transpose the previous vector and the transpose the result:

$$\Theta = \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \end{bmatrix} - 0.0025(\begin{bmatrix} 0.99 \\ 0.99 \\ -0.02 \\ -0.03 \end{bmatrix}^T \begin{bmatrix} 1 & 5 & 3 \\ 1 & 5 & 3 \\ 1 & 3 & 3 \\ 1 & 2 & 4 \end{bmatrix})^T$$

$$\Theta = \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \end{bmatrix} - 0.0025 \begin{bmatrix} 1.93 \\ 9.78 \\ 5.76 \end{bmatrix}$$

$$\Theta = \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \end{bmatrix} - \begin{bmatrix} 0.005 \\ 0.024 \\ 0.014 \end{bmatrix} = \begin{bmatrix} 0.495 \\ 0.476 \\ 0.486 \end{bmatrix}$$

4 Some basic statistics

5 Bonus: cost Logistic regression