Machine Learning 2016/2017: Written Assignment 2

Isabel Walter (worked with Carine Candel)

Deadline: 7 October 2016

1. (GRADED) This question is about vectorization, writing in vector form, of the update rule for multivariate linear regression.

a. $h_\theta(x) = [\theta_0 \quad \theta_1 \quad \cdots \quad \theta_n] \begin{bmatrix} x_0^{(i)} \\ x_1^{(i)} \\ \vdots \\ x_n^{(i)} \end{bmatrix}$

b. $J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta \left(\begin{bmatrix} x_0^{(i)} \\ x_1^{(i)} \\ \vdots \\ x_n^{(i)} \end{bmatrix} \right) - y^{(i)})$^2

c. $\frac{\partial J(\theta)}{\partial \theta} = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta \left(\begin{bmatrix} x_0^{(i)} \\ x_1^{(i)} \\ \vdots \\ x_n^{(i)} \end{bmatrix} \right) - y^{(i)} \right) * \begin{bmatrix} x_0^{(i)} \\ x_1^{(i)} \\ \vdots \\ x_n^{(i)} \end{bmatrix})$

d. $\begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(\left(h_\theta \left(\begin{bmatrix} x_0^{(i)} \\ x_1^{(i)} \\ \vdots \\ x_n^{(i)} \end{bmatrix} \right) - y^{(i)} \right) * \begin{bmatrix} x_0^{(i)} \\ x_1^{(i)} \\ \vdots \\ x_n^{(i)} \end{bmatrix} \right)$

2.

a.

X	y	freq	P(X=x, Y=y)
0	0	a	a/(a+b+c+d)
0	1	b	b/(a+b+c+d)
1	0	c	c/(a+b+c+d)
1	1	d	d/(a+b+c+d)

b. P(X = 0) = (a+b)/(a+b+c+d)

c. P(X = 1|Y = 0) = (P(X=1)∩P(Y=0))/(P(Y=0)) = (a/(a+b+c+d)/((a+c)/(a+b+c+d))

d. P(X = 1 ∪ Y = 0) = P(X=1) + P(Y=0) - P(X=1)∩P(Y=0)

e.

$$\boldsymbol{COV(X,Y) = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{n-1}}$$

3. (GRADED) We have the following data X:2, 5, 7, 7, 9, 25.

a. μ =(2 + 5 + 7 + 7 + 9 + 25)/6 = 9.1667
σ^2 = 1/5((2 – μ)^2 + (5 – μ)^2 + (7 – μ)^2 + (7 – μ)^2 + (9 – μ)^2 + (25 – μ)^2)) = 65.7667

b. P(x) = $\frac{1}{\sqrt{2\pi\sigma^2}} * e^{\frac{-(x-\mu)^2}{2\sigma^2}}$ fill in x = 20, μ = 9.1667 and σ^2. Gives: P(20) = 0.020

c. P(2) * P(5) * P(7) * P(7) * P(9) * P(25) = 0.03329 * 0.04311 * 0.04747* 0.04747 * 0.04918 * 0.007314= 1.1633 * 10^-9

d. P(8) = 0.04869 ..
P(25) is smaller (being 0.0073) → so bigger.

e.
MeanY = (4 + 4 + 5 +6 +8 +10)/ 6 = 6.1667
Cov(X,Y) = E((X – E(X))(Y – E(Y)) =

$$\boldsymbol{COV(X,Y) = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{n-1}}$$

Cov(X,Y) = 17.5667

f.
For covariance definition see e.
MSE:

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

Covariance: You have two different sets of numbers (X and Y) with for each number in X a corresponding number in Y (they form pairs). For each set you calculate the difference between the mean of the set and a certain number in that set (Note that the number you use for the X set corresponds to the number you use fort the Y set), then you multiply the outcomes. You do this for every X and Y pair and sum the results. After this you divide the whole thing by n – 1, n being the amount of samples (X and Y pairs).

MSE: You have a set y with certain values and a set h that forms the prediction of set y. So set h has a predicted value of each value in set y. Thus, naturally both sets have the same size. The MSE formula sums the squared differences between the two sets and then divides it by 2m. m being the size of the sets.

Both the MSE and the covariance calculate differences; MSE calculates the difference between two sets and the covariance calculates the difference between the mean of a set and each sample in that set for two different sets with corresponding numbers. However the MSE calculates a difference between sets and the covariance calculates differences within sets and then multiplies them.

Both the MSE and the covariance divide these differences by (a form of) the total size of the sets. However this is for both a different form (either 2m or n-1).

Both the MSE and the covariance use two sets of the same size.

4.
a. ??????
b/c The threshold number becomes too high, since for each feature the probability lowers (bcs the total probability remains 1). Therefore you need to decrease the threshold.