Machine Learning 2016-2017: Written Assignment 2

Question One:

This question is about vectorization, i.e. writing expressions in matrix-vector form. The goal is to vectorize the update rule for multivariate linear regression.

*a) Let ϑ be the parameter vector ϑ = (ϑ0 ϑ1 ... ϑn)T and let the i-th data vector be: x(i) = (x0 x1 ... xn)T where x0 = 1. What is the vectorial expression for the hypothesis function hϑ(x)?*

hϑ(x) = (ϑ0 ϑ1 ... ϑn)T \* (x0 x1 ... xn) = ϑT \* x(i)

*b) What is the vectorized expression for the cost function: J(ϑ) (still using the explicit summation over all training examples).*

J(ϑ) = 1/2m ∑ (ϑT \* x(i) - y(i))^2

*c) What is the vectorized expression for the gradient of the cost function, i.e. what is:*

*[ δJ(ϑ) / δϑ0 ]*

*[ δJ(ϑ) / δϑ1 ]*

*δJ(ϑ) / δϑ = [ . ]*

*[ . ]*

*[ δJ(ϑ) / δϑn ]*

*Again the explicit summantion over the data vectors from the learning set is allowed here.*

[ ∑ (ϑT \* x(i) - y(i)) \* x(i)0 ]

[ ∑ (ϑT \* x(i) - y(i)) \* x(i)1 ]

δJ(ϑ) / δϑ = 1/m [ . ]

[ . ]

[ . ]

[ ∑ (ϑT \* x(i) - y(i)) \* x(i)n ]

*d) What is the vectorized expression for the ϑ update rule in the gradient descent procedure.*

[ ∑ (ϑT \* x(i) - y(i)) \* x(i)0 ]

[ ∑ (ϑT \* x(i) - y(i)) \* x(i)1 ]

ϑj := ϑj + α/m [ . ]

[ . ]

[ . ]

[ ∑ (ϑT \* x(i) - y(i)) \* x(i)n ]

Question Three:

We assume the value 2, 5, 7, 7, 9, 25 are random values from a normal distribution.

*a) Estimate the mean, μ, and the variance, σ^2 of this normal distribution.*

μ = 1/m \* ∑ x(i) = 1/6 \* (2 + 5 + 7 + 7 + 9 + 25) = 9.17

σ^2 = 1/m \* ∑ ( x(i) - μ)^2 = 1/6 \* ((2 - 9.17)^2 + (5 - 9.17)^2 + (7 - 9.17)^2 \* 2 + (25 - 9.17)^2) = 54.80

*b) Let X ~ N(μ, σ^2) be a random variable. Calculate the probability density* ƒx(20)

general formula is:

ƒ(x; μ, σ^2) = 1/ σ√(2π) \* exp(-1/2 \* ((x - μ) / σ) ^ 2)

P(20) = 1/(√(2 π )√(54.80)) \* exp (-1/2 \* ((20 - 9.17) / 54.80)^2)

P(20) = 0.0528492

*c) Now consider six random variables X1, .., Xn. All independent of eachother and all identically and normally distributed with mean μ and variance σ^2 as calculated above. Let ƒX1 .. X6(x1, ..., x6) be the joint probability density function. Calculate ƒX1 .. X6(2, 5, 7, 7, 9, 25).*

P(2, 5, 7, 7, 9, 25) = ∏ P(x(i))  
 P(2) = 0.0534321

P(5) = 0.0537356

P(7) = 0.0538492

P(9) = 0.0538912   
 P(25) = 0.0516892

P(2, 5, 7, 7, 9, 25) = 2.3192 \* 10^-8

d) Is *ƒX1 .. X6(2, 5, 7, 7, 8, 9) larger or smaller then the probability density calculated above?*

Because P(25) is smaller then P(8) that means that the probability increases slightly, which means Is ƒX1 .. X6(2, 5, 7, 7, 8, 9) is larger.

*e) Now consider two random variables X and Y and six random samples of this multivariate distribution:*

x y

\_\_\_\_\_\_\_

2 4

5 4

7 5

7 6

9 8

25 10

*Estimate the covariance cov(X, Y).*

general formula is:

Cov(X, Y) = 1/m ∑[ (x(i) - μX) \* (y(i) - μY) ]

μX = mean of X : 9.17

μY = mean of Y = (4 + 4 + 5 + 6 + 8 + 10)/ 6 : 6.17

Cov(X, Y) = 1/6 \* ∑[ (x(i) - 9.17) \* (y(i) - 6.17) ]

(2−9.17) ∗ (4−6.17) = 15.5589

(5−9.17) ∗ (4−6.17) = 9.0489

(7−9.17) ∗ (5−6.17) = 2.5389

(7−9.17) ∗ (6−6.17) = 0.36890  
 (9 − 9.17) ∗ (8 − 6.17) = −0.3111

(25 − 9.17) ∗ (10 − 6.17) = 60.6289

Cov(X,Y) = 1/6 \* (15.5589 + 9.0489 +2.5389 +0.36890 + −0.3111 + 60.6289) = 14.6389

*f) Compare the defenition of the covariance with the mean squared error that is used in the cost function in linear regression. Are they related? Is there a difference? If so, what? Explain your answer.*

The Mean Squared Error is 1/2m ∑[ hϑ(x(i)) - y(i) ] ^ 2

If you split this you see this is taking the mean of ( hϑ(x(i)) - y(i) )^2

The Covariance function is 1/m ∑[ (x(i) - μX) \* (y(i) - μY) ]

If you split that you can also see that you calculate the mean of (x(i) - μX) \* (y(i) - μY)

Although they look quite similar there is quite a big difference, because with the Mean Squared Error you calculate the relationship between the real value and the expected value and with covariance you look at the relation between the x and y value.