HW4

Stav Naky and Oron Barazani

Question 1- (i)

The definition of correlation is the covariance divided by the product of the standard deviations.

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y}$$

So, if we prove that the covariance between random independent X and Y is zero, hence their correlation is zero.

The definition of covariance is the expected product of two variables subtracted by the product of their expected values:

Cov(X,Y)=$E[XY]-E[X]E[Y]$

So, if we prove that E[XY] = E[X]E[Y] then the covariance (hence correlation) is zero.

The definition of expected value of the product is:

$$E[XY] = \sum_x \sum_y xy \cdot P(X = x, Y = y)$$

And, If X and Y are random independent variables, then by definition their joint probability equals the product of their probabilities.

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$$

So:

$$E[XY] = \sum_x \sum_y xy \cdot P(X = x) \cdot P(Y = y)$$

$$E[XY] = \sum_x P(X = x) \cdot \sum_y P(Y = y)$$

Which equals to E[X]·E[Y]. Thus, the correlation between random independent variables is zero.

Question 1-(ii)

No, because correlation only implies that there is a **linear** relationship between two variables. If there is no linear relationship (not correlated), it doesn’t necessarily mean that there is independence, because there can be another type of relationship between the two variables, for example- the activity of an enzyme and the temperature. It’s a parabolic relationship- so no correlation, but of course they’re not independent variables.

Question 2

(i) $X \text{ and } X^2$ are always independent- **False**. X^2 will always be explained by X, hence dependent. Given X= [1,4,5,8,10], X^2 = [1,16,25,64,100] - we can see that there’s a relationship between X and X^2 by the covariance.

(ii) $X \text{ and } X^2$ are never correlated- **False**. They **CAN** have a correlation of zero, but it’s not always the case. An example of when a correlation is zero is when the mean equals 0. But when the values are more positive or more negative, then the correlation will be higher. For example-

X=[-10,0,10] X^2=[100,0,100] - correlation will be zero.

X=[10,50,100] X^2 = [100, 2500, 10000] - correlation will be positive.

(iii) $X \text{ and } X^2$ are always correlated- **False**. As I have shown before, it can vary.

(iv) $X \text{ and } X^2$ are never independent - **True.** X^2 can always be explained by X hence is never independent.

<u>Question 3</u>

I'll break it down to two proofs: a. symmetrical and b. semi-positive definite.

a.Symmetry

In order to prove symmetry, we need to prove $C = C^T$.

We know from the given $C = X^T \cdot X$ that $C^T = (X^T \cdot X)^T$

So:

$$C^T = (X^T)^T \cdot X^T = X \cdot X^T = C$$

So I proved $C = C^T$ thus symmetrical.

b. Semi-positive definite

Let v be a non-zero column vector, and v^T be a non-zero row vector. To prove that C is a semi-positive definite, we need to prove that:

$$v^T C v \geq 0.$$

The definition of C is $X \cdot X^T$, hence the expression becomes:

$v^T \cdot X \cdot X^T v$

And since $(v^T \cdot X) = (X^T \cdot v)^T$:

$(X^T \cdot v)^T \cdot (X^T \cdot v) = ||X^T \cdot v||^2$

And since the squared norm of a vector is always non-negative, we can conclude that C is a semi-positive definite.

Question 4

(i) Given that $X = \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 3 & 0 \\ 5 & 6 \\ 6 & 6 \\ 7 & 6 \end{bmatrix}$, its covariance matrix is: $C = X^T X = \begin{bmatrix} 124 & 108 \\ 108 & 108 \end{bmatrix}$

(ii) First we'll find the eigenvalues of the matrix, using the characteristic polynomial:

$$|\lambda I - C| = \begin{vmatrix} \lambda - 124 & 108 \\ 108 & \lambda - 108 \end{vmatrix} = (\lambda - 124)(\lambda - 108) - 108^2 =$$
$$= \lambda^2 - 232\lambda + 1728$$

We can find its roots:

$$\lambda_1 = 224.2959 \qquad \lambda_2 = 7.7041$$

Which are the eigenvalues of C:

$$D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} 224.2959 & 0 \\ 0 & 7.7041 \end{bmatrix}$$

Now we'll find the matching eigenvectors. We need to find the solution to the homogeneous equation (x is 2x1 vector):

$$(C - \lambda I)x = 0$$

For each λ.
First:

$$\begin{bmatrix} 124 - 224.2959 & 108 \\ 108 & 108 - 224.2959 \end{bmatrix} x = 0$$
$$\begin{bmatrix} -100.2959 & 108 \\ 108 & -116.2959 \end{bmatrix} x = 0$$

The rows are dependent, so we remain with one dimension in the null space solution:

$$\begin{bmatrix} -100.2959 & 108 \\ 0 & 0 \end{bmatrix} x = 0$$
$$x_1 = \begin{bmatrix} -0.7328 \\ -0.6805 \end{bmatrix}$$

Same for the other vector:

$$\begin{bmatrix} 124 - 7.7041 & 108 \\ 108 & 108 - 7.7041 \end{bmatrix} x = 0$$
$$\begin{bmatrix} 116.2959 & 108 \\ 108 & 100.2959 \end{bmatrix} x = 0$$
$$\begin{bmatrix} 116.2959 & 108 \\ 0 & 0 \end{bmatrix} x = 0$$
$$x_2 = \begin{bmatrix} 0.6805 \\ -0.7328 \end{bmatrix}$$

So in general, the eigen vectors are the columns of U the diagonalize matrix of C. moreover, because that U columns form an orthonormal basis is equivalent to that U is unitary matrix, so $U^T = U^{-1}$ and the final eigen-decomposition is:

$$C = UDU^{-1} = UDU^T$$
$$= \begin{bmatrix} 0.6805 & -0.7328 \\ -0.7328 & -0.6805 \end{bmatrix} \begin{bmatrix} 7.7041 & 0 \\ 0 & 224.2959 \end{bmatrix} \begin{bmatrix} 0.6805 & -0.7328 \\ -0.7328 & -0.6805 \end{bmatrix}$$

(iii) The main component is the eigen-vector of C that is matching the biggest eigen-value. In our case it is:

$$x_1 = \begin{bmatrix} -0.7328 \\ -0.6805 \end{bmatrix}$$

Most of the variance of data is spread across this axis.

(iv) The main eigen-value is the biggest eigen-value and represents the amount of variability (variance) in that axis (eigen-vector).

(v) To project new data point on the main component:

$$x_7 x_1 = [2,1] \begin{bmatrix} -0.7328 \\ -0.6805 \end{bmatrix} = -2.146$$

This is the projection of the 2D sample on the main component axis

Q5 - 0ff part

(a) $W^{(1)} = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}$

$W^{(2)} = \begin{bmatrix} w_{31} \\ w_{32} \end{bmatrix}$

(b) First, lets define v_1, v_2:

$v_1 = w_{11} \cdot 1 + w_{12} \cdot x$

$v_2 = w_{21} \cdot 1 + w_{22} \cdot x$

Now let's define o_1, o_2:

$o_1 = \varphi(v_1) = \varphi(w_{11} + w_{12} \cdot x)$

$o_2 = \varphi(v_2) = \varphi(w_{21} + w_{22} \cdot x)$

So, $\hat{y}(x) = w_{31} \cdot o_1 + w_{32} \cdot o_2 = w_{31} \cdot \varphi(w_{11} + w_{12} \cdot x) + w_{32} \cdot \varphi(w_{21} + w_{22} \cdot x)$.

(c) The function $y(x)$ is:

$y = \begin{cases} \max(0, \frac{-x-1}{2}) & x < 0 \\ \max(0, x-1) & x \geq 0 \end{cases}$

And $\varphi(x) = \max(0, x)$, so: $\varphi(\frac{-x-1}{2}) = \max(0, \frac{-x-1}{2})$

$\varphi(x-1) = \max(0, x-1)$

$y = \begin{cases} \varphi(\frac{-x-1}{2}), & x < 0 \\ \varphi(x-1), & x \geq 0 \end{cases}$

When will $\hat{y} = y$?

When:

$w_{31} \cdot \varphi(w_{11} + w_{12} \cdot x) + w_{32} \cdot \varphi(w_{21} + w_{22} \cdot x) = \begin{cases} \varphi(\frac{-x-1}{2}), & x < 0 \\ \varphi(x-1), & x \geq 0 \end{cases}$

(c) continued:

With set of weights:

$w_{11} = -\frac{1}{2}$, $w_{12} = -\frac{1}{2}$

$w_{21} = 1$, $w_{22} = -1$

$w_{31} = 1$, $w_{32} = 1$

Wight:

for $x \geq 0$:

$\hat{y} = 1 \cdot \varphi(-\frac{1}{2} - \frac{x}{2}) + 1 \cdot \varphi(-1+x)$

$\varphi(x-1) = \max(x-1, 0) = 0$

$\hat{y} = \varphi(\frac{-x-1}{2}) = y(x)$

for $x < 0$:

$\hat{y} = 1 \cdot \varphi(\frac{-x-1}{2}) + 1 \cdot \varphi(x-1)$

$\varphi(\frac{-x-1}{2}) = \max(\frac{-x-1}{2}, 0) = 0$

$\hat{y} = y^2$

Q5 - ors part

(d) $L = (y - \hat{y}(x))^2$

$\vec{\nabla} L$ will have 6 elements, one for each weight.

$$\nabla L = \left[\frac{\partial L}{\partial w_{11}}, \frac{\partial L}{\partial w_{12}}, \dots \frac{\partial L}{\partial w_{32}}\right]$$

A common term that can help us would be:

$\frac{\partial L}{\partial \hat{y}} = \delta = 2 \cdot (\hat{y}(x) - y(x))$, and we will apply the chain rule each time.

$$\frac{\partial L}{\partial w_{11}} = \delta \cdot \frac{\partial \hat{y}(x)}{\partial w_{11}} = \delta \cdot \frac{\partial (w_{31} \cdot \varphi(w_{11} + w_{12}x) + w_{32} \cdot \varphi(w_{21} + w_{22}x))}{\partial w_{11}} =$$

$$= \delta \cdot (w_{31} \cdot \varphi'(w_{11} + w_{12}x) \cdot 1 + w_{32} \cdot 0) = \delta \cdot w_{31} \cdot \varphi(w_{11} + w_{12}x)$$

$$\frac{\partial L}{\partial w_{12}} = \delta \cdot \frac{\partial (w_{31} \cdot \varphi(w_{11} + w_{12}x) + w_{32} \cdot \varphi(w_{21} + w_{22}x))}{\partial w_{12}} = \delta \cdot w_{31} \cdot \varphi'(w_{11} + w_{12}x) \cdot x$$

$$\frac{\partial L}{\partial w_{21}} = \delta \cdot \frac{\partial (w_{31} \cdot \varphi(w_{11} + w_{12}x) + w_{32} \cdot \varphi(w_{21} + w_{22}x))}{\partial w_{21}} = \delta \cdot w_{32} \cdot \varphi'(w_{21} + w_{22}x)$$

$$\frac{\partial L}{\partial w_{22}} = \delta \cdot w_{32} \cdot \varphi'(w_{21} + w_{22}x) \cdot x$$

$$\frac{\partial L}{\partial w_{31}} = \delta \cdot \varphi(w_{11} + w_{12}x)$$

$$\frac{\partial L}{\partial w_{32}} = \delta \cdot \varphi(w_{21} + w_{22}x)$$

$$\boxed{\text{So } \nabla L = [\delta \cdot w_{31} \cdot \varphi'(w_{11} + w_{12}x), \delta \cdot w_{31} \cdot \varphi'(w_{11} + w_{12}x) \cdot x, \delta \cdot w_{32} \cdot \varphi'(w_{21} + w_{22}x), \delta \cdot w_{32} \cdot \varphi'(w_{21} + w_{22}x) \cdot x, \delta \cdot \varphi(w_{11} + w_{12}x), \delta \cdot \varphi(w_{21} + w_{22}x)]}$$

ORG part – Q5

(e) $w_{11}' = -\eta \cdot \delta \cdot w_{31} \cdot \varphi'(w_{11} + w_{12}x) + w_{11}$

$w_{12}' = -\eta \cdot \delta \cdot w_{31} \cdot \varphi'(w_{11} + w_{12}x) \cdot x + w_{12}$

$w_{21}' = -\eta \cdot \delta \cdot w_{32} \cdot \varphi'(w_{21} + w_{22}x) + w_{21}$

$w_{22}' = -\eta \cdot \delta \cdot w_{32} \cdot \varphi'(w_{21} + w_{22}x) \cdot x + w_{22}$

$w_{31}' = -\eta \cdot \delta \cdot \varphi(w_{11} + w_{12}x) + w_{31}$

$w_{32}' = -\eta \cdot \delta \cdot \varphi(w_{21} + w_{22}x) + w_{32}$

(f) Given: $x=0, y=0, w_{11}=1, w_{12}=1, w_{21}=-1, w_{22}=-1, w_{31}=1, w_{32}=1$

$L = (y-\hat{y})^2$

$y = 0$

$\hat{y} = 1 \cdot \varphi(1+1 \cdot 0) + 1 \cdot \varphi(-1-1 \cdot 0) = 1 \cdot 1 + 1 \cdot (0) = 1$

~~$L = (0-0)^2 = 0$~~

$L = (0-1)^2 = 1$