hw1

April 10, 2020

1 Homework 1

Name: **Juan Vila**

1.1 Question 1

Which of the following matrices are positive semi-definite and hence valid covariance matrices? ### Section A We know that the condition for proof that a matrix A is semi defintive positive is that for a non zero vector v we have the following result $v^{T}Av >= 0 $

$$\begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$$

Solving this multiplication we got that $ a(a+b) + b(a+b) = (a+b)^2 $ which is positive for any value of a and b, then definite positive, then semi definite positive. ### Section B We know that the condition for proof that a matrix A is semi defintive positive is that for a non zero vector v we have the following result $v^{T}Av >= 0 $

$$\begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$$

Solving this multiplication we got that $ 2ab + b^2 = b(2a+b) $ which is positive if $b > 0$ and $2a+b > 0$ or can be negative if $b < 0$ and $2a+b < 0$ which implies that is not semi-definite positive. ### Section C We know that the condition for proof that a matrix A is semi defintive positive is that for a non zero vector v we have the following result $v^{T}Av >= 0 $

$$\begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$$

Solving this multiplication we got that $ -ab + ab + b^2 = b^2 $ which is positive for any value of a and b, then definite positive, then semi definite positive. However, it's not symetric, this implies that is not a covariance matrix ### Section D We know that the condition for proof that a matrix A is semi defintive positive is that for a non zero vector v we have the following result $v^{T}Av >= 0 $

$$\begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$$

Solving this multiplication we got that $ a^2 -ab - ab + b^2 = (a-b)^2 $ which is positive for any value of a and b, then definite positive, then semi definite positive.

1.2 Question 2

1.2.1 Section A

One way to show this proposition is by contradiction. Using the normal equation $X^T X\hat{\theta} = X^T Y$ and assuming that there is another vector of parameters $\hat{\alpha}$ that minimize the lost function we could write the model as $Y = \hat{\alpha}X$ then if we replace into the normal equation we got that $X^T X\hat{\theta} = X^T X\hat{\alpha}$. Then if it's invertible implies that $\hat{\alpha} = \hat{\theta}$, meaning that $\hat{\theta}$ is unique.

1.2.2 Section B

In this part we are going to argue as $p > n$ implies that $rank(X) = n$ which implies that $rank(X^T X) = n$, but the dimensions of $X^T X$ is pxp, which means that are at least $p-n$ rows and columns that are LD, which implies that when we try to solve the normal equations we have more variables than equations which means that the system of equations have infinites solutions.

1.3 Question 3

1.3.1 Section A

Bias

$$E\left[\hat{\mu}\right] - \mu = E\left[\frac{1}{m}\sum_{i=1}^{m} z_i\right] - \mu$$

$$E\left[\hat{\mu}\right] - \mu = \frac{1}{m}\sum_{i=1}^{m} E\left[z_i\right] - \mu$$

$$E\left[\hat{\mu}\right] - \mu = \frac{1}{m}\sum_{i=1}^{m} \mu - \mu$$

$$E\left[\hat{\mu}\right] - \mu = \frac{1}{m} m\mu - \mu$$

$$E\left[\hat{\mu}\right] - \mu = \mu - \mu$$

$$E\left[\hat{\mu}\right] - \mu = 0$$

Variance

$$Var\left[\hat{\mu}\right] = Var\left[\frac{1}{m}\sum_{i=1}^{m} z_i\right]$$

$$Var\left[\hat{\mu}\right] = \frac{1}{m^2}\sum_{i=1}^{m} Var\left[z_i\right]$$

$$Var\left[\hat{\mu}\right] = \frac{1}{m^2}\sum_{i=1}^{m} \sigma^2$$

$$Var\left[\hat{\mu}\right] = \frac{1}{m^2} m\sigma^2$$

$$Var\left[\hat{\mu}\right] = \frac{\sigma^2}{m}$$

MSE

$$MSE = Bias^2 + Variance$$

$$MSE = \frac{\sigma^2}{m}$$

1.3.2 Section B

Bias

$$E\left[\hat{\mu_0}\right] - \mu_0 = E\left[0\right] - \mu_0$$

$$E\left[\hat{\mu_0}\right] - \mu_0 = -\mu_0$$

Variance

$$Var\left[\hat{\mu_0}\right] = Var\left[0\right]$$

$$Var\left[\hat{\mu_0}\right] = 0$$

MSE

$$MSE = Bias^2 + Variance$$

$$MSE = -\mu_0$$

1.3.3 Section C

We know that the estimator is $\hat{\mu}_{\lambda,m} = \lambda\hat{\mu}_m + (1-\lambda)\hat{\mu}_0$, the the bias is #### Bias

$$E\left[\hat{\mu}_{\lambda,m}\right] - \mu_{\lambda,m} = E\left[\lambda\hat{\mu}_m + (1-\lambda)\hat{\mu}_0\right] - \lambda\mu_m + (1-\lambda)\mu_0$$

$$E\left[\hat{\mu}_{\lambda,m}\right] - \mu_{\lambda,m} = \lambda E\left[\hat{\mu}_m\right] + (1-\lambda)E\left[\hat{\mu}_0\right] - \lambda\mu_m - (1-\lambda)\mu_0$$

$$E\left[\hat{\mu}_{\lambda,m}\right] - \mu_{\lambda,m} = \lambda\mu_m - \lambda\mu_m - (1-\lambda)\mu_0$$

$$E\left[\hat{\mu}_{\lambda,m}\right] - \mu_{\lambda,m} = -(1-\lambda)\mu_0$$

Variance

$$Var\left[\hat{\mu}_{\lambda,m}\right] = Var\left[\lambda\hat{\mu}_m + (1-\lambda)\hat{\mu}_0\right]$$

$$Var\left[\hat{\mu}_{\lambda,m}\right] = \lambda^2 Var\left[\hat{\mu}_m\right] + (1-\lambda)^2 Var\left[\hat{\mu}_0\right]$$

$$Var\left[\hat{\mu}_{\lambda,m}\right] = \lambda^2\frac{\sigma^2}{m}$$

MSE

$$MSE = Bias^2 + Variance$$

$$MSE = (-(1-\lambda)\mu_0)^2 + \lambda^2\frac{\sigma^2}{m}$$

$$MSE = (1-2\lambda+\lambda^2)\mu_0^2 + \lambda^2\frac{\sigma^2}{m}$$

Now that we have an expresion for MSE, we can minimize λ, for doing this we can take the derivate. The only thing that we have to worry is that the function is convex, since if it is convex, the optimization result is a minimium. This can be ensure because λ square have positive signs (the variance is always positive and μ_0 is squared), meaning that the result is minimium.

$$\frac{\partial MSE}{\partial\lambda} = (-2+2\lambda)\mu_0^2 + 2\lambda\frac{\sigma^2}{m}$$

$$(-2+2\lambda)\mu_0^2 + 2\lambda\frac{\sigma^2}{m} = 0$$

$$\lambda\mu_0^2 + \lambda\frac{\sigma^2}{m} = \mu_0^2$$

$$\lambda^* = \frac{\mu_0^2}{\mu_0^2 + \frac{\sigma^2}{m}}$$

1.4 Question 4

1.4.1 Section A

We know that the covariance is

$$COV\left[\hat{Y}\right] = E\left[\left[Y - E\left[\hat{Y}\right]\right]\left[Y - E\left[\hat{Y}\right]\right]^T\right]$$

For calculating the covariance we need to calculate the expected value of Y, this is:

$$Y = X\theta + \epsilon$$

$$E\left[Y\right] = E\left[X\theta\right] + E\left[\epsilon\right]$$

$$E\left[Y\right] = \theta E\left[X\right] + E\left[\epsilon\right]$$

Using the facts that the mean of θ and ϵ is zero, we replace an get the result that expected value of Y is zero. Now if we replace this fact into the covarinace we get the following result.

$$COV\left[\hat{Y}\right] = E\left[[Y]\,[Y]^T\right]$$

Now if we replace by the process that create Y, we have the following result:

$$COV\left[\hat{Y}\right] = E\left[[X\theta + \epsilon]\,[X\theta + \epsilon]^T\right]$$

$$COV\left[\hat{Y}\right] = E\left[\theta X X^T \theta^T + 2X^T\theta^T\epsilon + \epsilon^T\epsilon\right]$$

Using the fact that constants we can get out of the expectation operator, ϵ and σ are independent, the covariance equal to zero, and replacing the variances of ϵ and σ we get:

$$COV\left[\hat{Y}\right] = \sigma_\theta^2 X X^T + \sigma_\epsilon^2$$

1.4.2 Section B

Since the p vector are orthonormal, impliest that $X^T X = I$, then we can argue that :

$$B = XX^T$$

$$XB = XX^T X$$

$$XB = X$$

Using this result we can modify the covariance result in the following wat

$$COV\left[\hat{Y}\right] X = \sigma_\theta^2 I_p X X^T X + \sigma_\epsilon^2 X$$

$$COV\left[\hat{Y}\right] X = (\sigma_\theta^2 I_p + \sigma_\epsilon^2) X$$

Then we can see that X is the eigenvector for the covariance matrix and the eigenvalues is $(\sigma_\theta^2 I_p + \sigma_\epsilon^2)$

1.4.3 Section C

[113]:
```
import numpy as np
import matplotlib.pyplot as plt
M = 10000
n = 32
p = 2
sig_t = 0.1
sig_e = 0.01;
#t = (0:(n-1))'/n;
#X = np.ones((n,p))
#for f in range(p-2):
#    X[:,f+2] = np.kron(np.ones(2^f,1),np.kron(np.transpose([1,-1]),np.ones(n/
↪2^(f+1),1)))
```

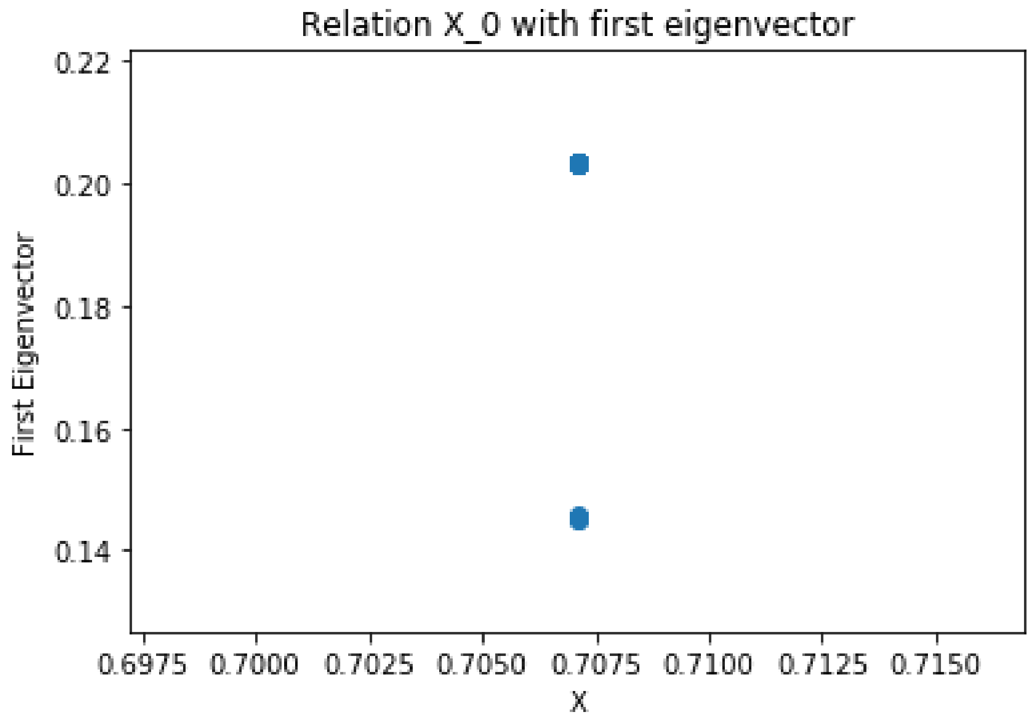
```
X = np.ones((n, p))
for f in range(p - 1):
    X[:, f + 1] = np.kron(np.ones(pow(2, f)), np.kron(np.transpose([1, -1]), np.
 ↪ones(n//pow(2, f+1))))

def normalize_rows(x: np.ndarray):
    """
    function that normalizes each row of the matrix x to have unit length.

    Args:
     ``x``: A numpy matrix of shape (n, m)

    Returns:
     ``x``: The normalized (by row) numpy matrix.
    """
    return x/np.linalg.norm(x, ord=2, axis=1, keepdims=True)

X = normalize_rows(X)
S = np.zeros(n)

for m in range(M-1):
    theta = np.random.randn(p,1)*sig_t
    y = np.dot(X,theta) + np.random.randn(n,1)*sig_e
    S = S + np.dot(y,np.transpose(y))
S = S/M
```

[115]:
```
eigenvals = np.linalg.eigvals(S)
```

[116]:
```
eigenvectors = np.linalg.eig(S)
```

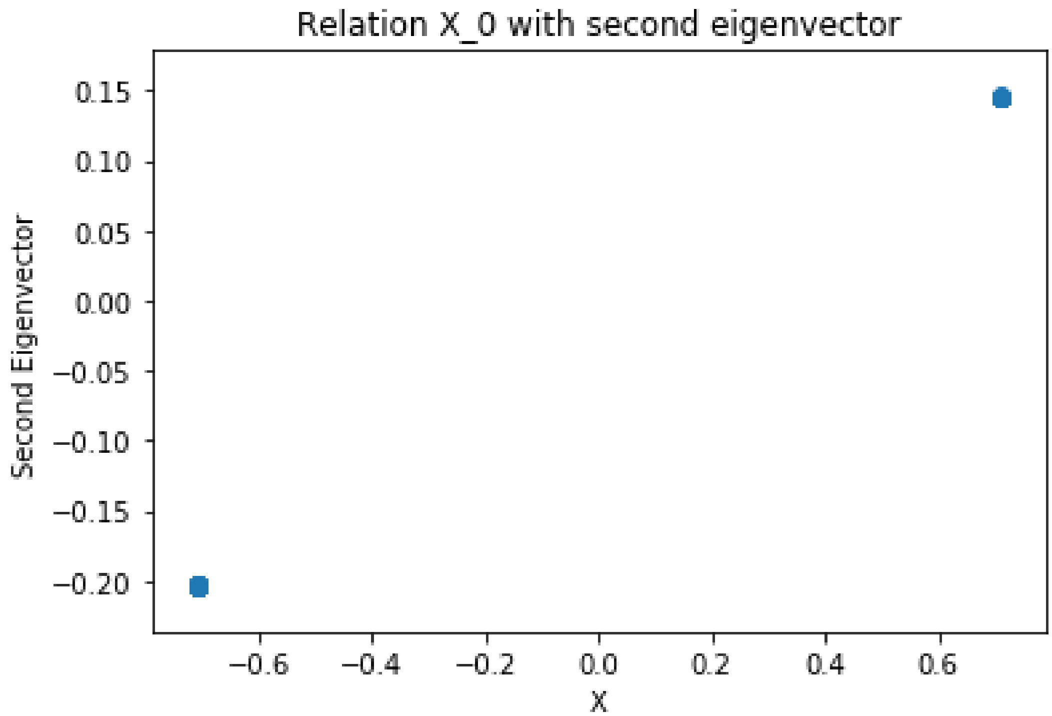
[133]:
```
plt.scatter(X[:,0],eigenvectors[1][:,0])
plt.title('Relation X_0 with first eigenvector')
plt.xlabel('X')
plt.ylabel('First Eigenvector')
```

[133]:
```
Text(0, 0.5, 'First Eigenvector')
```



[134]:
```
plt.scatter(X[:,1],eigenvectors[1][:,1])
plt.title('Relation X_1 with second eigenvector')
plt.xlabel('X')
plt.ylabel('Second Eigenvector')
```

[134]: Text(0, 0.5, 'Second Eigenvector')



We can see that in the case of the first and second eigenvectors we have a close relations with X, in the case of the scond one, the realtion is basically lineal. Which is relateed to the principal component direction.