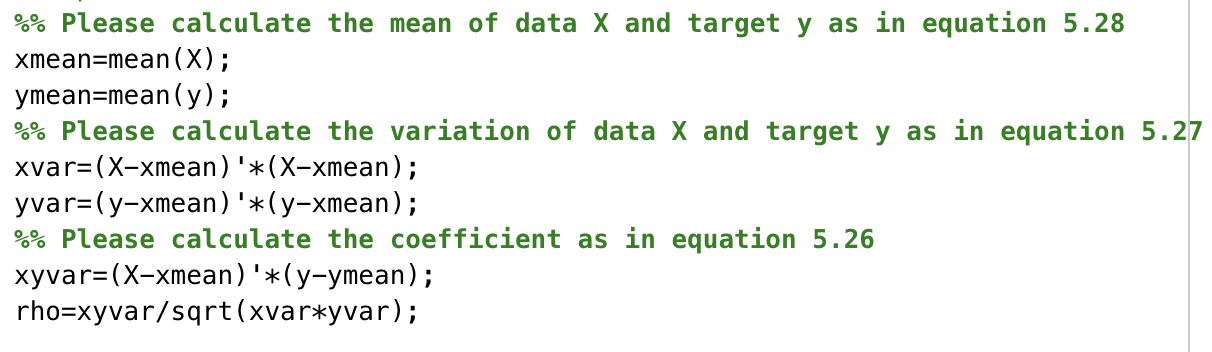
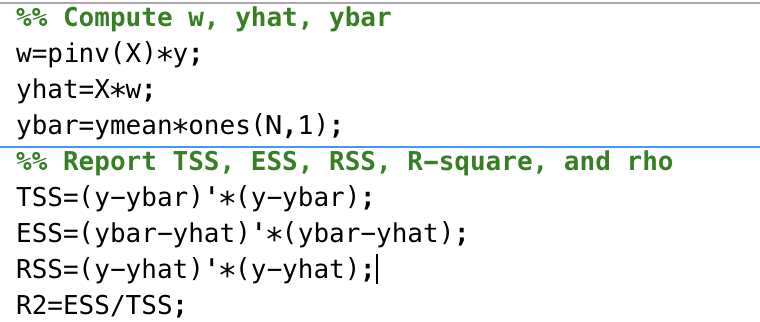
**Problem 3a**



Use the mean function in MATLAB to get the mean values (“xmean” and “ymean” which are scalar) of X (vector) and y (vector).

By the equation of variance and , we get the values of xvar, yvar and xyvar.

By definition of correlation coefficient, we get “rho=xyvar/sqrt(xvar\*yvar)”, so rho=0.7632.



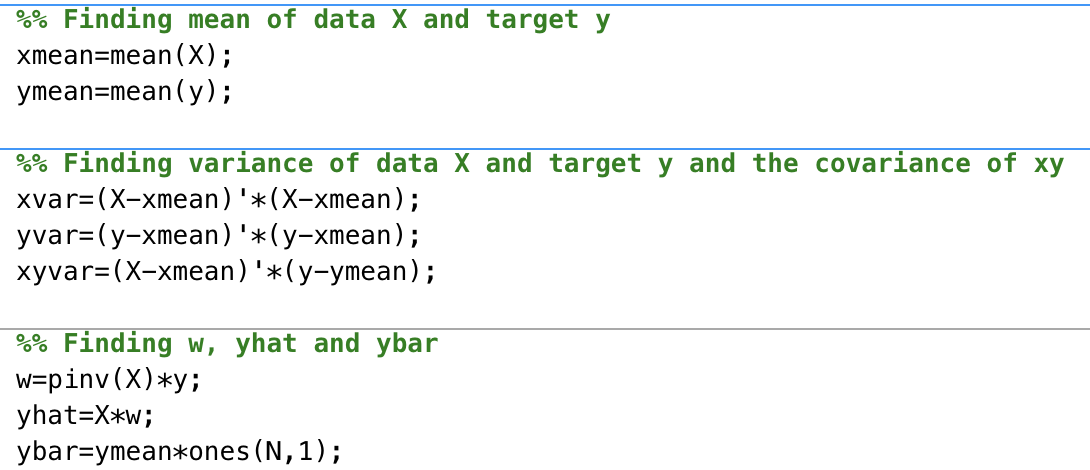
By the equation of the desired weight vector: , we get the value of w.

By the equation of the model prediction: , we get the value of yhat.

Because ymean is a scalar and ybar should be a vector, ybar = ymean\*ones(N,1).

By the definition of TSS, ESS, RSS and coefficient of determination we get the value of them.

**Problem 3b**



Use the mean function in MATLAB to get the mean values (“xmean” which is a 1x3 matrix and “ymean” is a scalar ) of X (matrix) and y (vector).

By the equation of variance and , we get the values of xvar, yvar and xyvar.

By definition of correlation coefficient, we get “rho=xyvar/sqrt(xvar\*yvar)”, so rho=0.7632.

A screenshot of a computer code

Description automatically generated

By the equation of the desired weight vector: , we get the value of w.

By the equation of the model prediction: , we get the value of yhat.

Because ymean is a scalar and ybar should be a vector, ybar = ymean\*ones(N,1).

By the definition of TSS, ESS, RSS and coefficient of determination we get the value of them.

**Problem 4**

As a result of the two variables being dependent (the eigenvalues of XX^T being close to 0 and having strong correlation coefficient), we should employ Ridge Regression rather than a simple linear regression.

Changing noise level

When experimenting with different values of **a**, the greater the value of **a**, the more severely **x2** deviates from **x1**, the smaller the correlation coefficient ρ (less correlate), and the larger the eigenvalues, the more they diverge from 0. In conclusion, the data gets well-conditioned with greater **a**.

i.e.

|  |  |  |  |
| --- | --- | --- | --- |
|  | a=1 | a=10 | a=100 |
| ρ | 0.9443 | 0.6001 | 0.4741 |
| eigenvalues | 0.241, 185.387 | 7.234, 618.5430 | 17.535, 25517.541 |

Changing step size

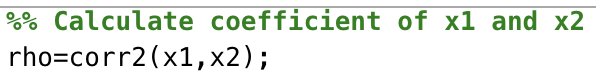
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| lambda | 0 | 0.1 | 0.2 | 0.4 | 0.8 |
| R-square | 0.9847 | 0.9783 | 0.9703 | 0.9587 | 0.9461 |
| w0 | 0.0582 | 0.2376 | 0.2979 | 0.3311 | 0.3313 |
| w1 | 2.2423 | 1.7650 | 1.5312 | 1.2975 | 1.1074 |
| w2 | -0.2357 | 0.1268 | 0.3118 | 0.5036 | 0.6652 |

It is clear that R-square declines as lambda increases, indicating that accuracy also declines with lambda increases.

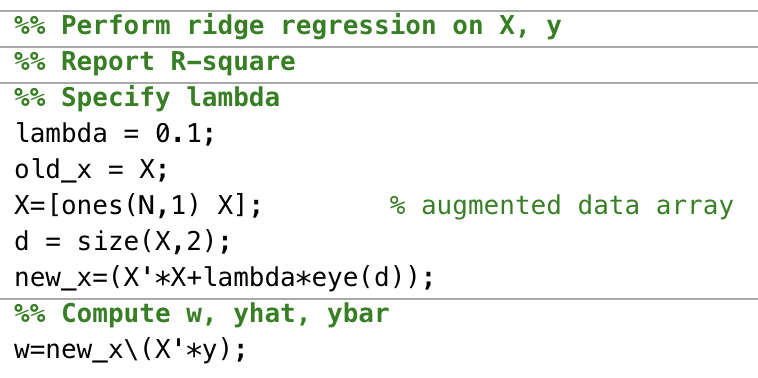
However, when lambda is raised, w0, w1, and w2 tend to remain steady.

As a result, we may draw the conclusion that an ill-conditioned problem's solution can be made more stable at the expense of accuracy.

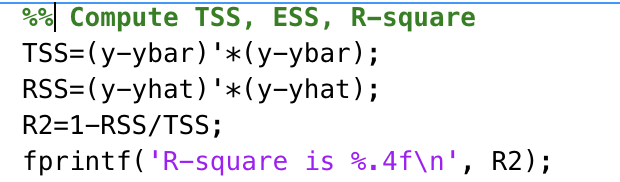
Coding



By definition of correlation coefficient, we get “rho=xyvar/sqrt(xvar\*yvar)”, but in there we use “corr2” function instead.



By the equation , we get w=new\_x\(X’\*y)



By the definition of TSS, RSS and coefficient of determination we get the value of them.