Vu Xuan Kim Cuong (Stephen)
1000646 – ESD Class of 2016

Question 2(d).

- Since the covariance matrix is just a diagonal matrix, it means that there is no correlation between the dimensions, and thus we can easily generate the vector such that each dimension follows univariate normal with mean 0 and variance

$$\sigma^2 = \frac{1.0}{i + 1.0} \; for \; i \; from \; 0 \; to \; 47$$

Question 3(d).

- Similarly, the covariance matrix is an identity matrix, which means we can generate the vector by generating 48 univariate normal random variable, with mean 0 and variance

$$\sigma^2 = 1.0$$

Question 6.

(c).
$Average\ MSE_{\lambda=e^{-30},n=100} = 31.2287495998,\ Average\ MSE_{\lambda=5,n=100} = 18.9139455322$

(d). There is a trade-off between $\lambda = e^{-30}$ and $\lambda = 5$: for $\lambda = 5$ the mean square error is smaller, but there is a bias. However, at low sample number, it is better to choose $\lambda = 5$

(f).
$Average\ MSE_{\lambda=e^{-30},n=500} = 16.1942046138,\ Average\ MSE_{\lambda=5,n=500} = 15.6105976969$

Now, the average MSE is comparable to each other, in this case it is better to choose $\lambda = e^{-30}$ since it has lower bias and reasonably low average.

Question 7.

(b). $A_1 =$ 30.2268800861 , V_1 =17.9145200882

(f). $A_2 =$ 30.1838397052 , V_2 =3.05221508022

Therefore, doing a 10-repetition, 5-fold cross-validation is better since it gives smaller variance.