2017 11 04 - Pset 06 - Q1

a) Consider mean of test error vs k

In k-fold cross validation, training data is divided into k folds. If we follow leave-one-out scheme, we will have (k-1) sets to learn parameter & k^{th} set to validate

Since k such trials are done implies that all k models learnt will be from data belonging to independent distribution.

For mean, we can comment that mean is a representative of the set of k-points sampled.

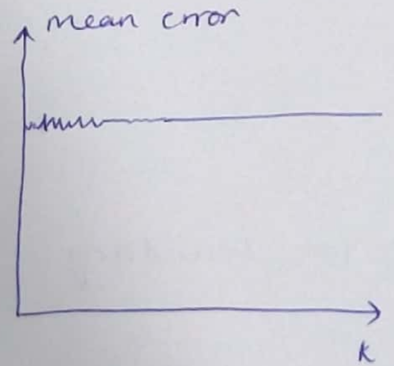
Now, consider the noise $(N(\mu, \sigma^2))$.

The outliers will not affect the mean of small sets ($k \sim$ order 100) to great extent if σ is sufficiently small ($\sigma \simeq 1$)

$\therefore$ Mean will stay nearly constant & may have slight jitters for k ($\because$ at extreme case $k=1$, mean = element value)

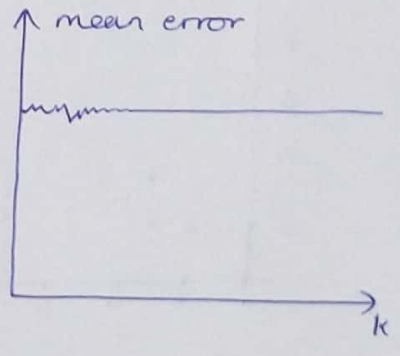
$\therefore$ Graph :

i)



sample = 100

ii)



sample = 1000

b) Variance is the measure of how far does a set of data move away from its eigen vectors.

Now, larger the value of k, smaller the size of individual set. Hence, smaller dimension of covariance matrix.

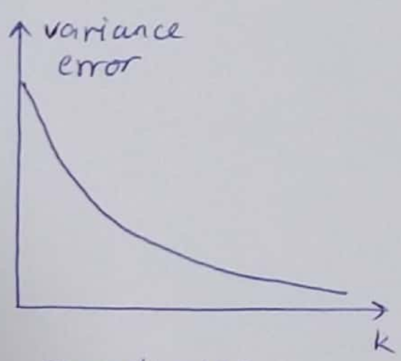
So, less information is available.

∴ Smaller variance with increasing k.

∴ For k = order 100, the variance error curve decays slower than for k = order 1000.

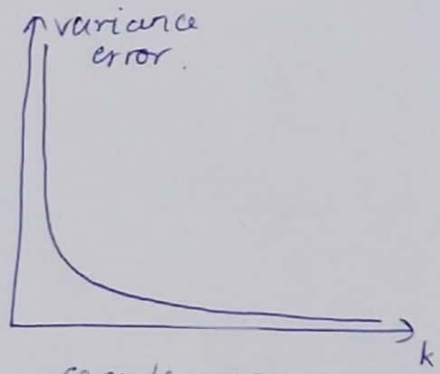
∴ Graph :

i)



sample = 100

ii)



sample = 1000