

Mean and variance of K-fold cross validation.

If we have K value to be too high, say a condition where each group has only 1 element, then we would not have enough training data to build a good predictive model. The consequence would be poor generalization and would work well only for the validation set.

If K was to be too low, there wouldn't be enough cross validation sets which would lead to overfitting.

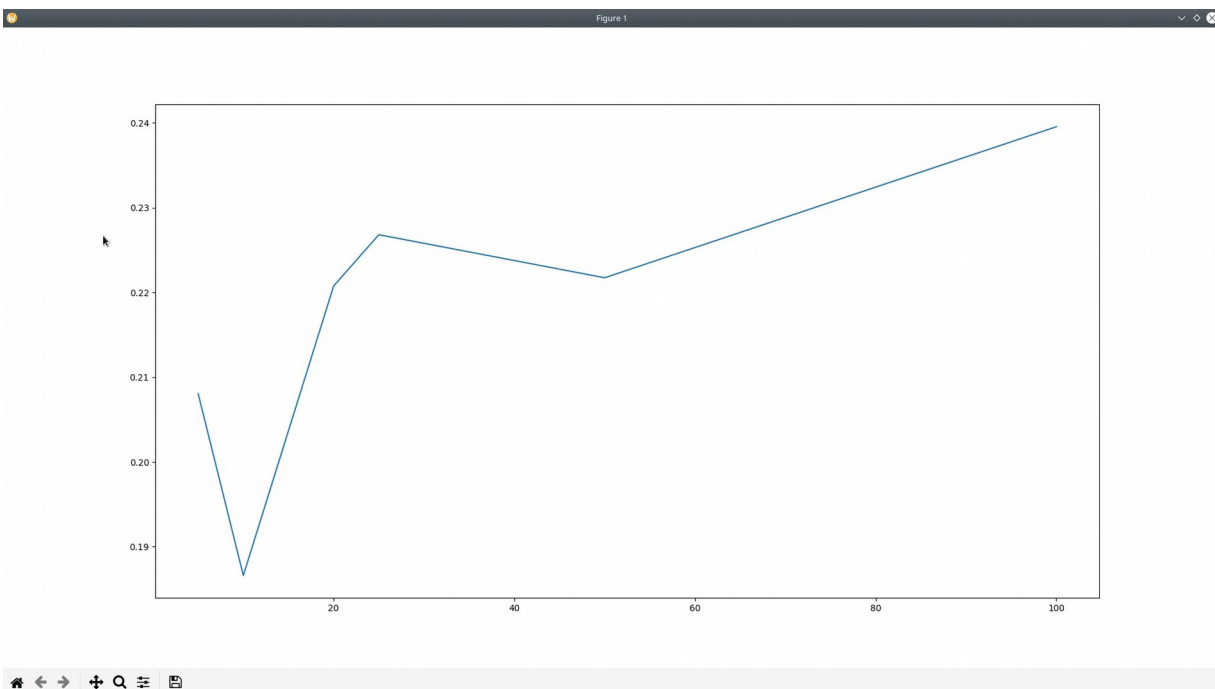
The ideal K would lie somewhere in between.

To speak of the influence on mean and variance -

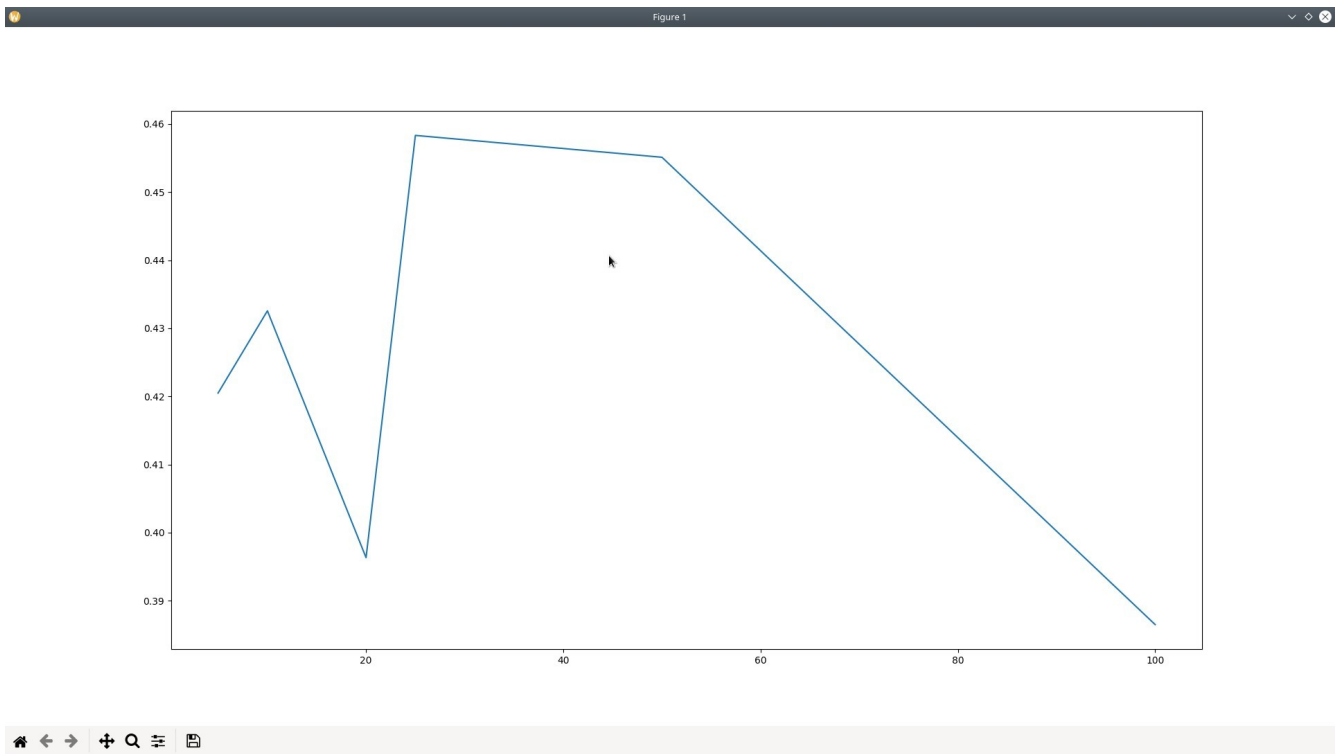
When K is low, the variance would be high.

When K is high, the variance would be low as there is a higher probability of finding a training set from K-1, which is closer to the validation set.

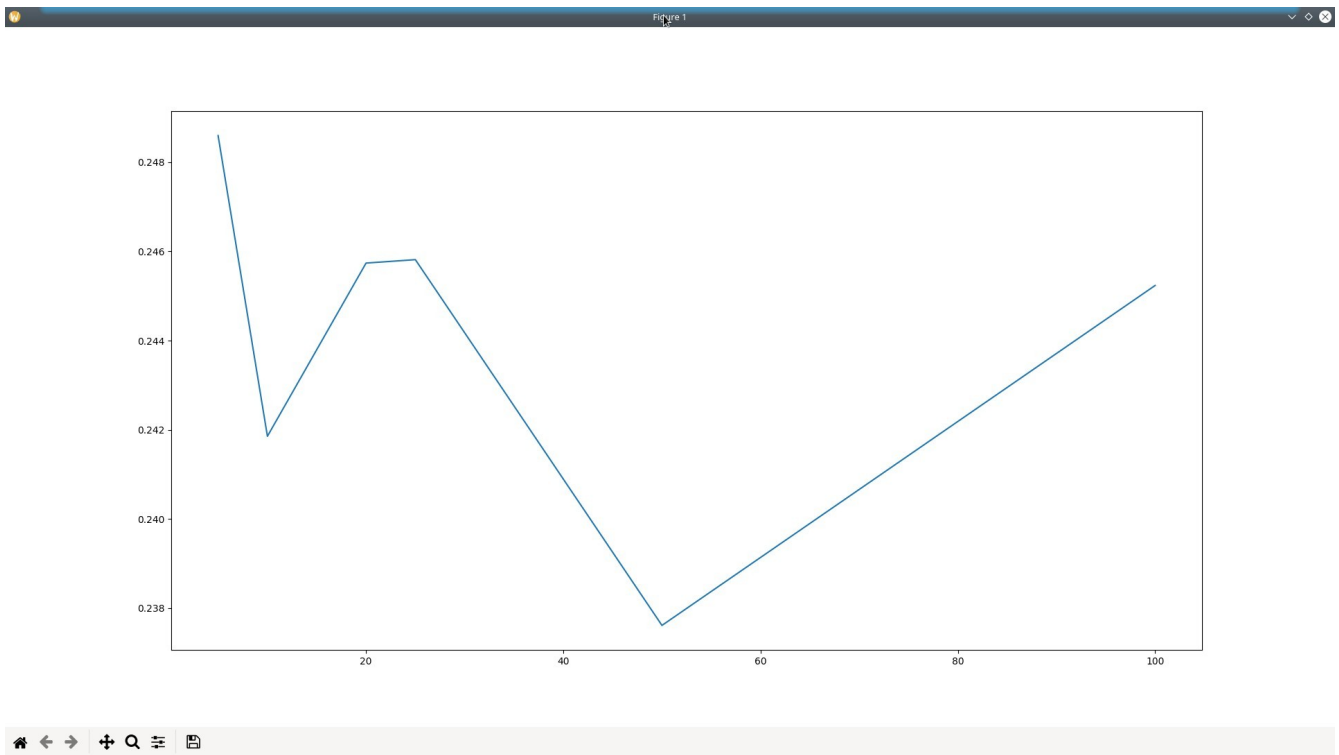
Case 1 Means



Case 1 Variances



Case 2 Means ::



Case 2 Vars ::

