

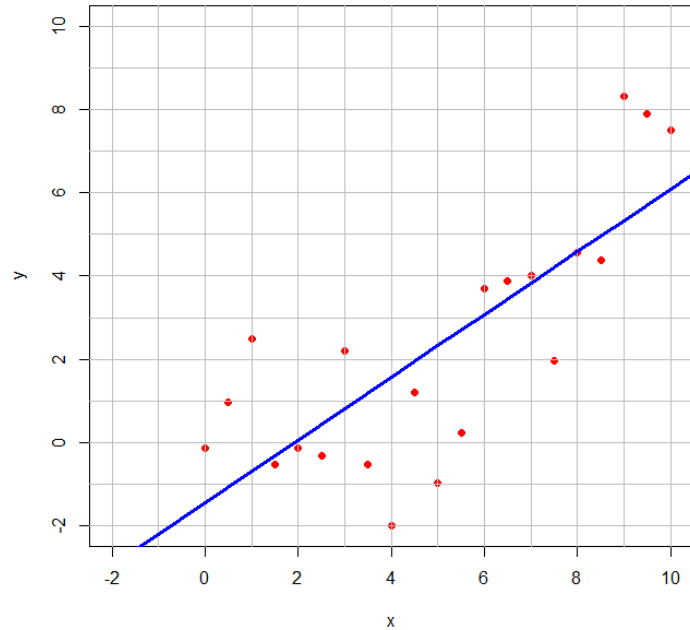
# Statistical Machine Learning

Day 7 – Bias and Variance in Linear Regression

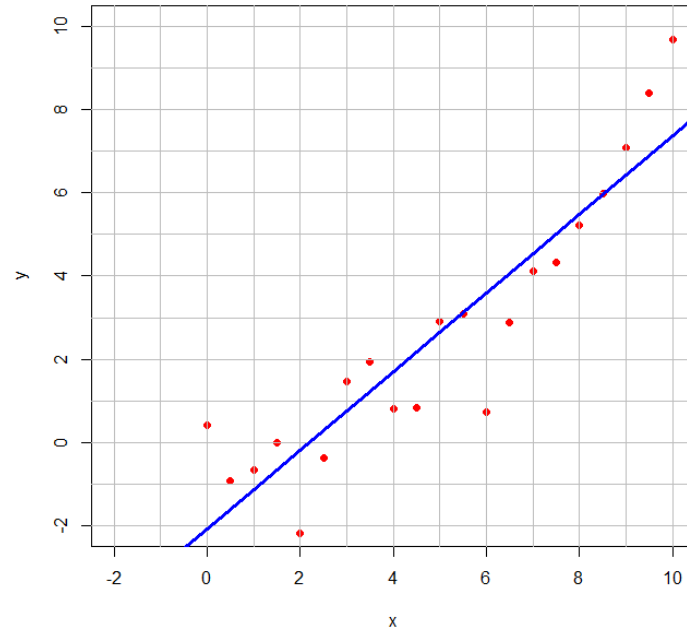
# Variance of Estimate over training sets

Linear estimates of  $f(x)$  from three different training sets

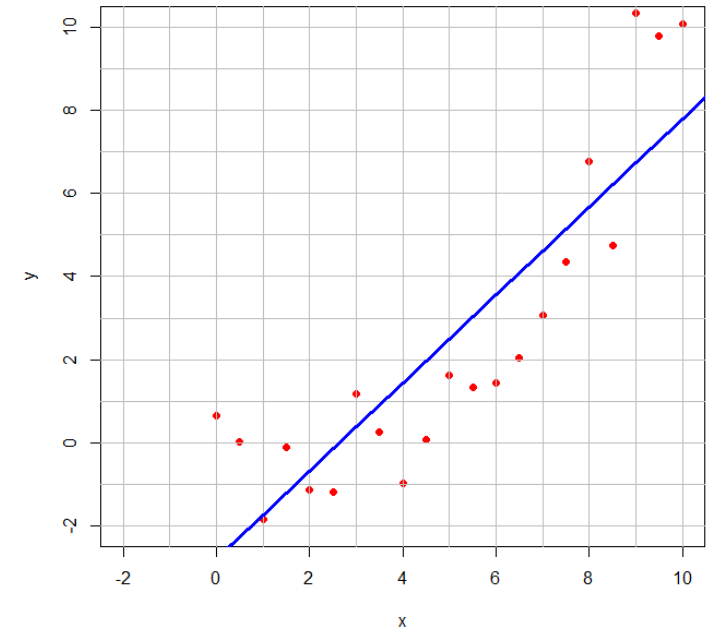
Training set of size 21 and linear model  
sample 1



Training set of size 21 and linear model  
sample 2



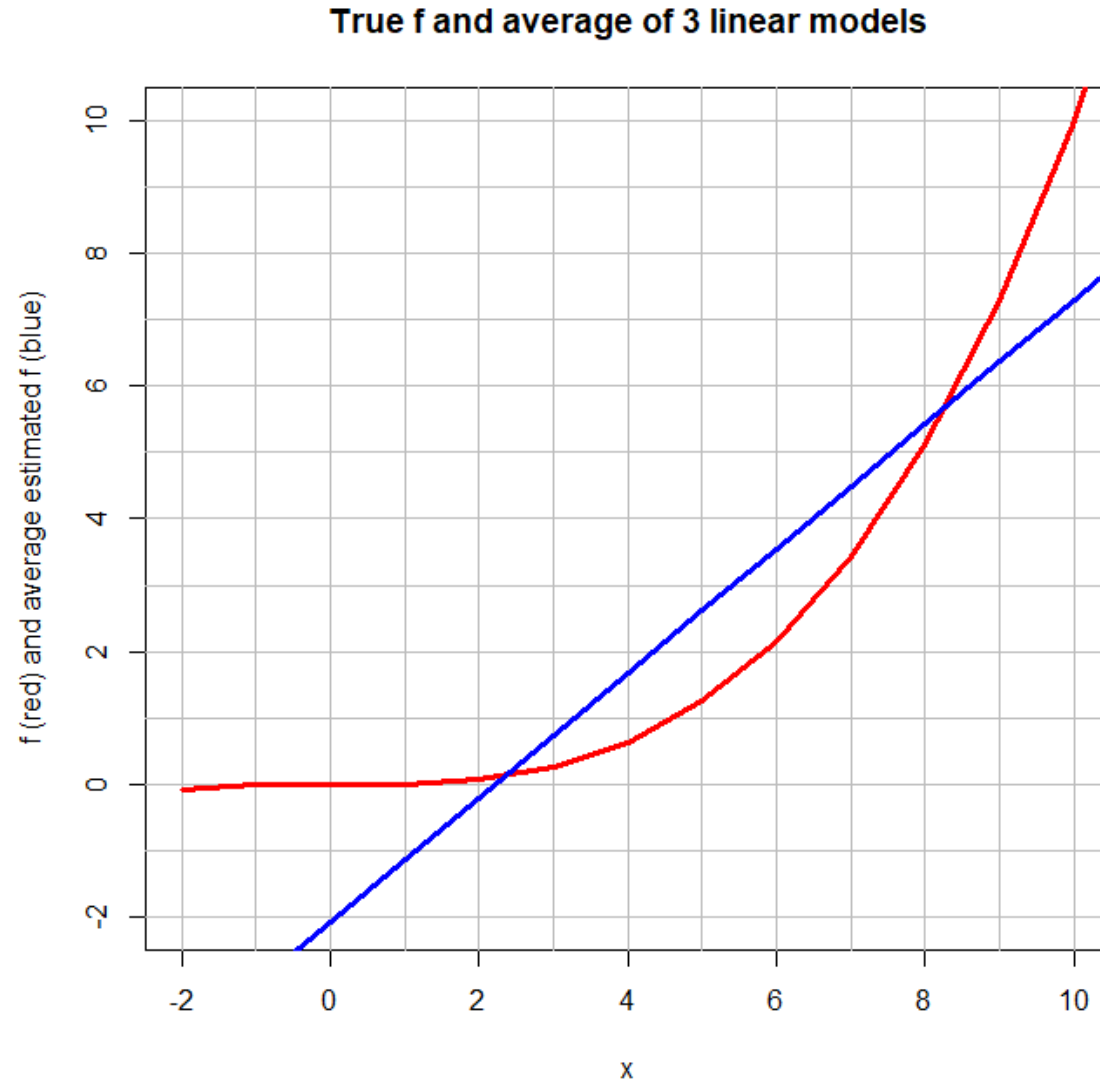
Training set of size 21 and linear model  
sample 3



# Bias between true f and average of linear estimates

Average squared bias over  $[0, 10]$ :

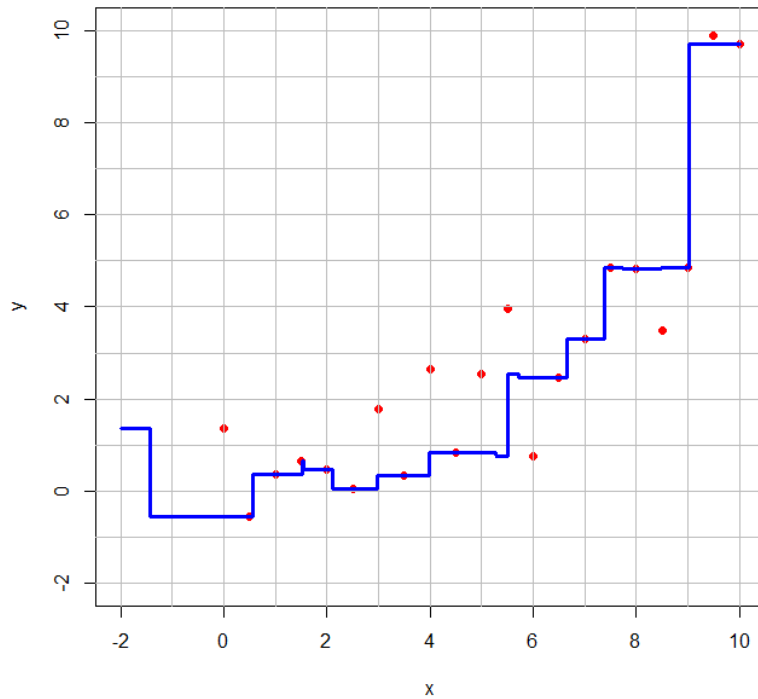
1.83



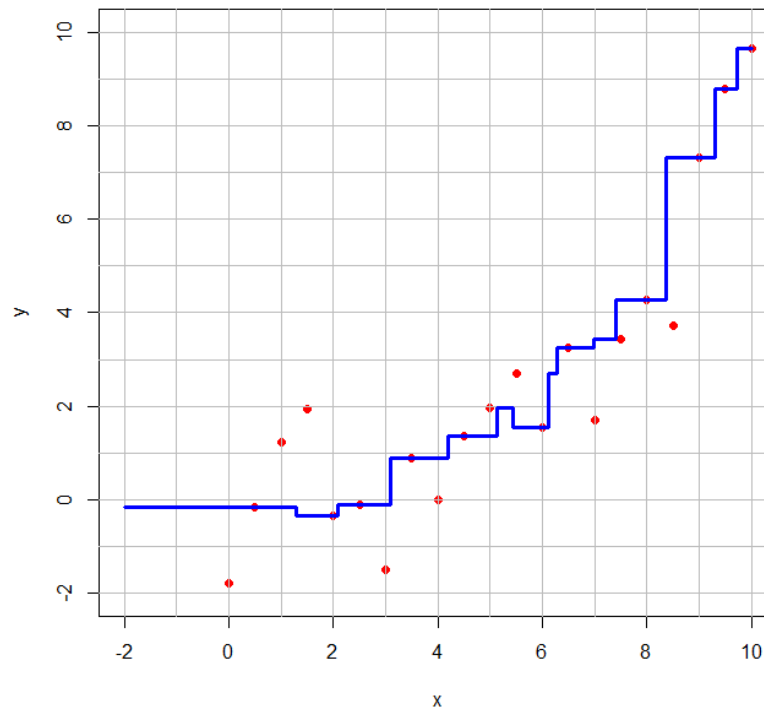
# Variance of Estimate over training sets

kNN,  $k=1$  estimates of  $f(x)$  from three different training sets

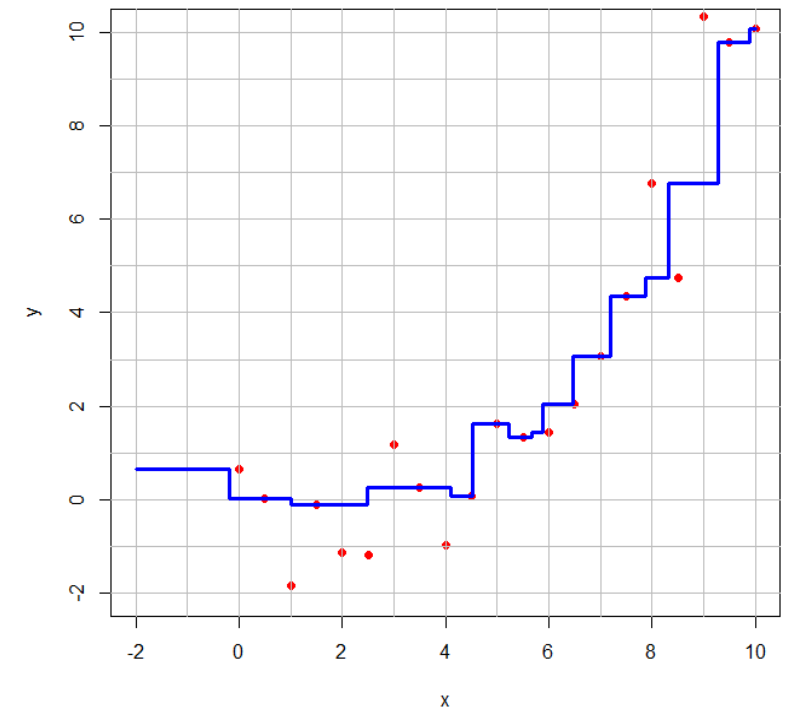
Training set of size 21 and kNN,  $k=1$   
sample 1



Training set of size 21 and kNN,  $k=1$   
sample 2



Training set of size 21 and kNN,  $k=1$   
sample 3



# Bias between true $f$ and kNN, $k=1$

Average squared bias  
over  $[0, 10]$ :

0.56

