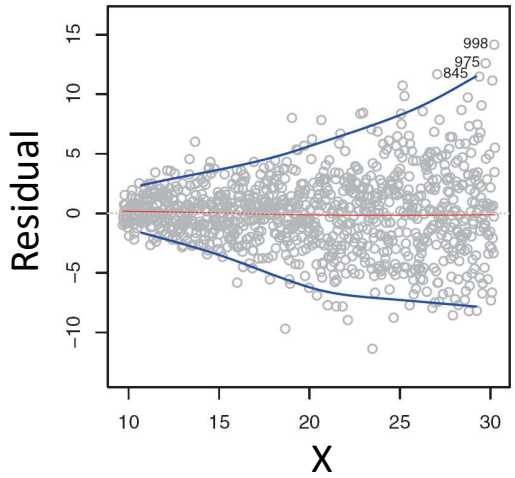
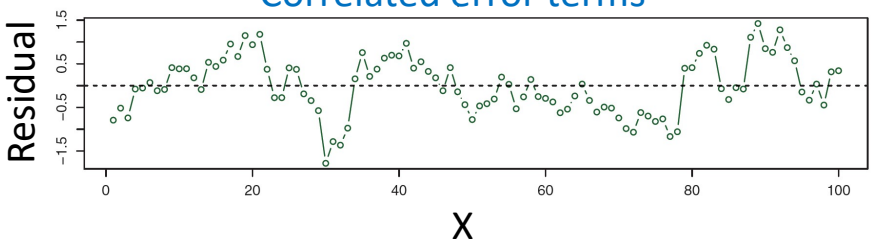
Notation and definitions

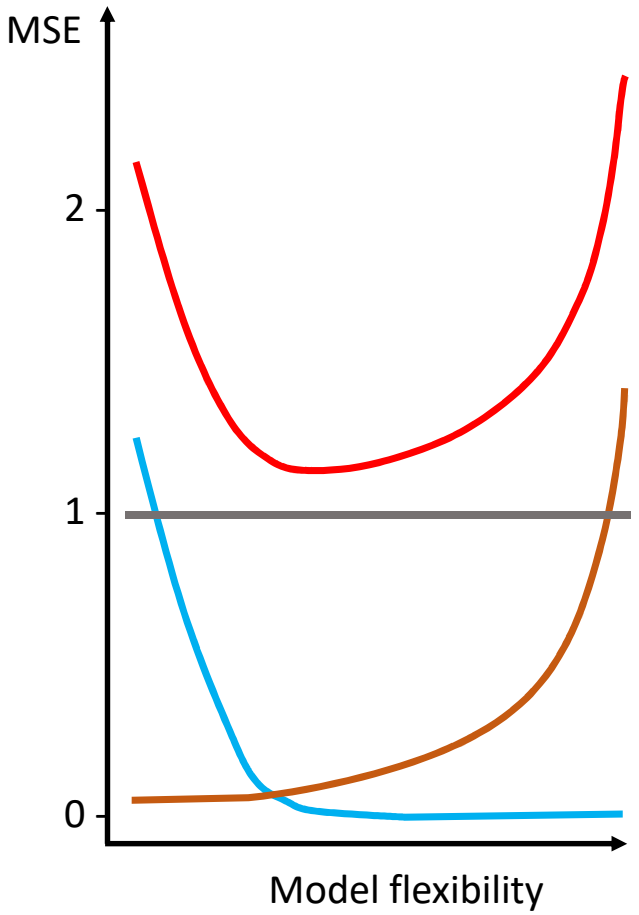
Non-constant variance (heteroscedasticity)



Correlated error terms



The Bias Variance trade-off

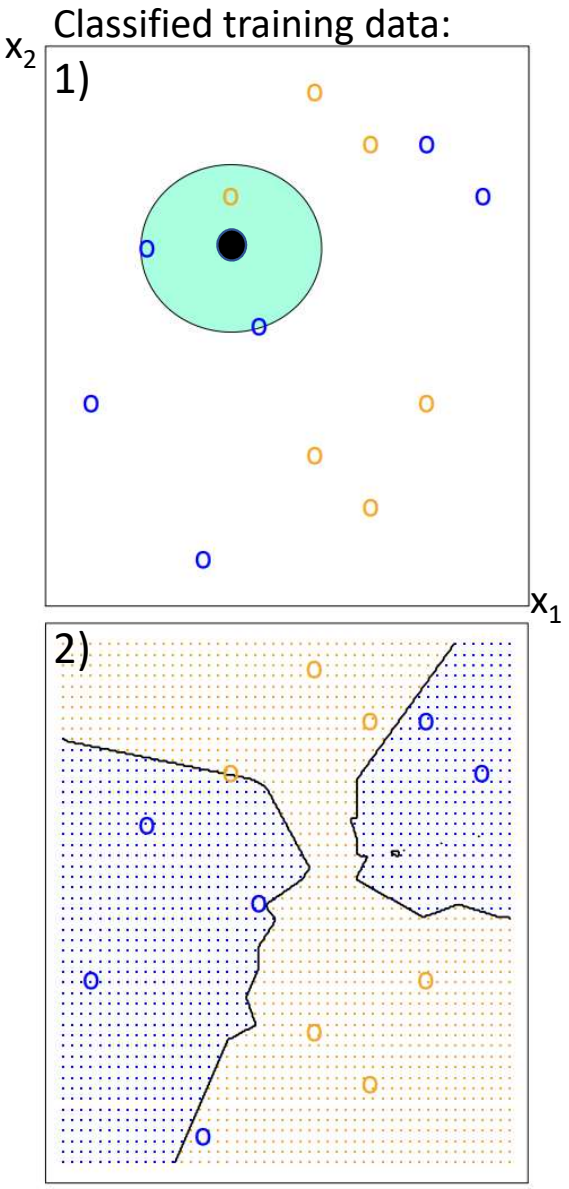


Quiz question

True or False:

A fitted model with more predictors will necessarily have a lower Training Set Error than a model with fewer predictors.

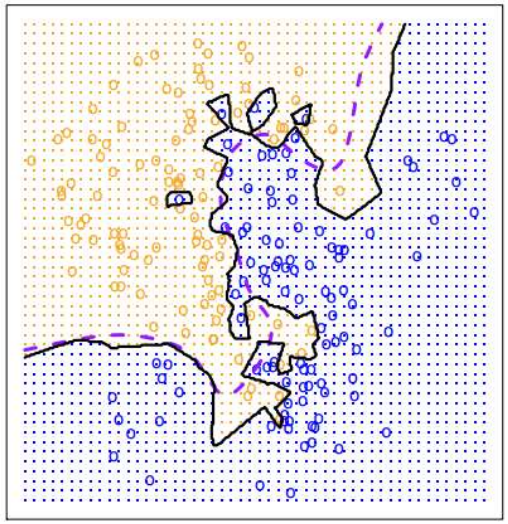
k-Nearest Neighbor: kNN

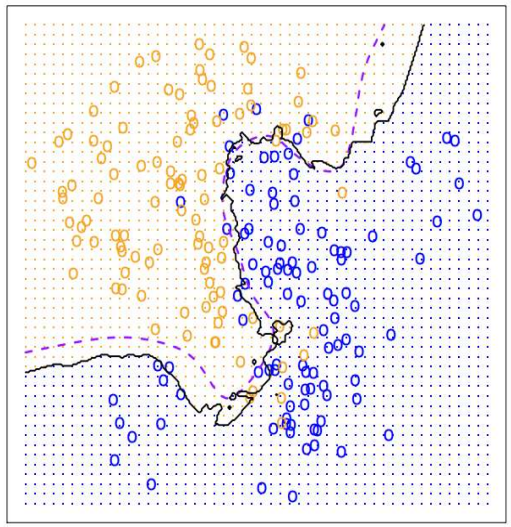


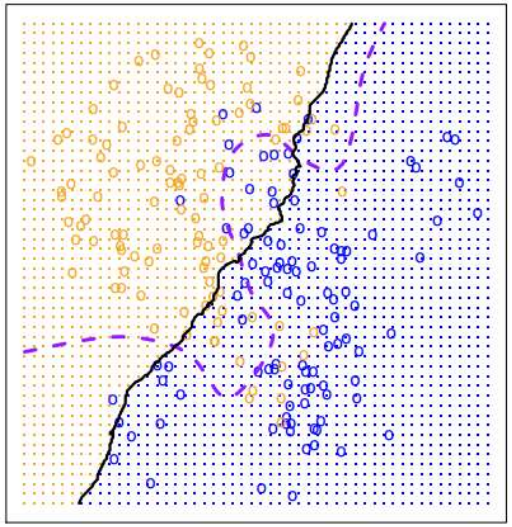
Figures from: An introduction to statistical learning, Hastie et al., 2021.

Quiz question

A nearest neighbor classifier was used here. What answer is correct?

1) 

2) 

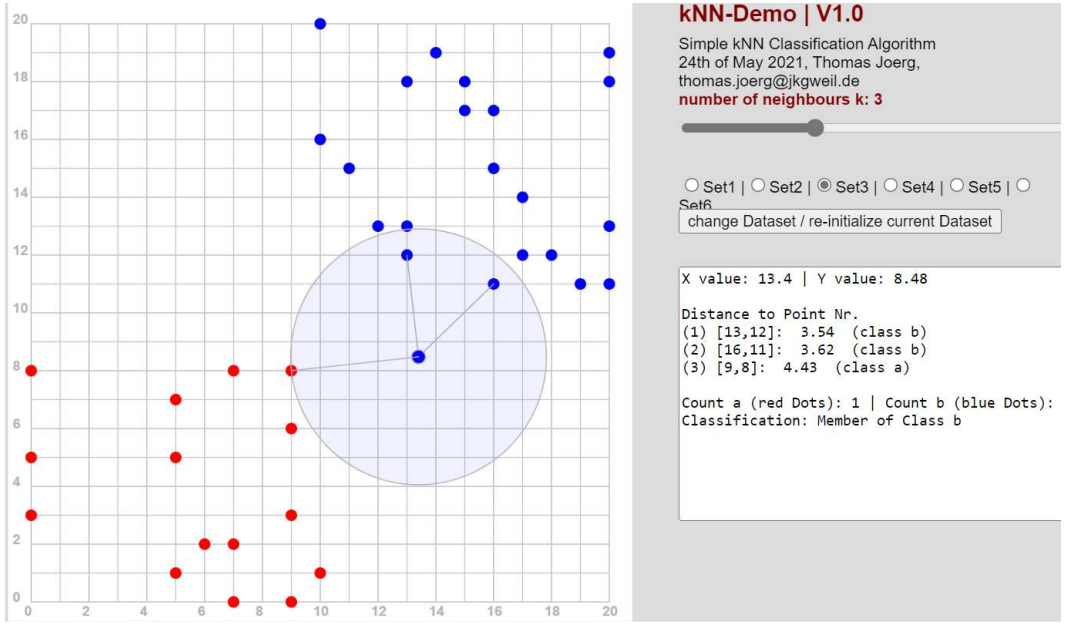
3) 

a)	$k_1 = 1$,	$k_2 = 100$,	$k_3 = 10$
b)	$k_1 = 10$,	$k_2 = 100$,	$k_3 = 1$
c)	$k_1 = 1$,	$k_2 = 10$,	$k_3 = 100$
d)	$k_1 = 100$,	$k_2 = 10$,	$k_3 = 1$

Figures from: An introduction to statistical learning, Hastie et al., 2021.

kNN simulator and programming exercise

https://iludis.de/kNNDemo/index.html



Run the kNN Jupyter Notebook in this weeks folder

Quiz question

While doing a homework assignment, you fit a Linear Model to your data set. You are thinking about changing the Linear Model to a Quadratic one. Which of the following is most likely true?

a) Using the quadratic model will decrease your irreducible error

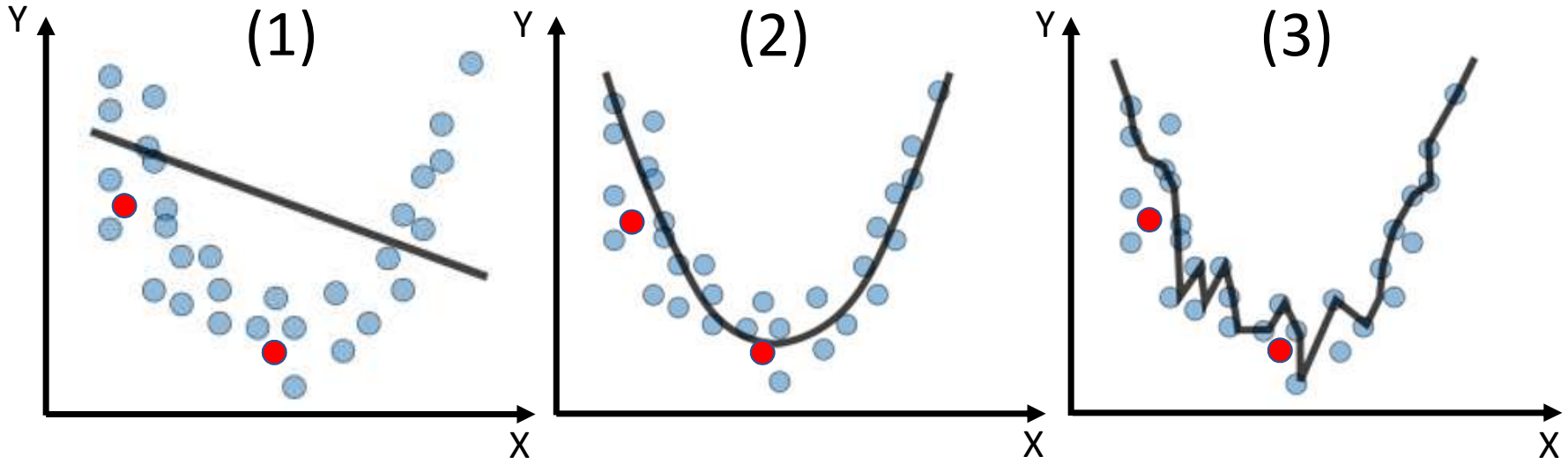
b) Using the quadratic model will decrease the bias of your model

c) Using the quadratic model will decrease the variance of your model

d) Using the quadratic model will decrease your reducible error

Quiz question

The black training data was fitted as shown below in order to predict the red test data points. What statements are correct?



a) Plot (1) shows a high variance
b) Plot (1) shows a high bias
c) Plot (3) shows a high bias
d) Plot (3) shows the phenomena of overfitting, i.e. the used model is too complicated

Quiz question

Consider the following figures.

1) Which one is related to a high bias issue?

2) Which ones is related to a high variance issue?