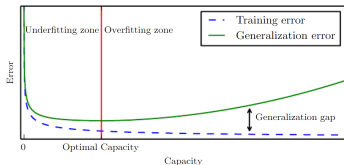


# Introduction to Machine Learning

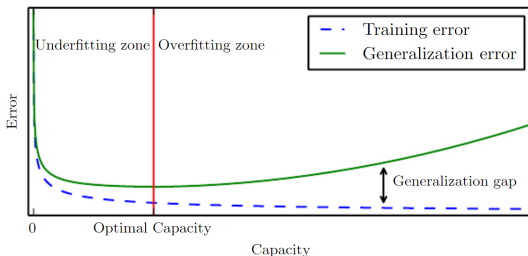
## Capacity & Overfitting



### Learning goals

- Know that the capacity of a hypothesis space impacts generalization
- Know that low capacity carries the risk of underfitting
- Know that too high capacity carries the risk of overfitting

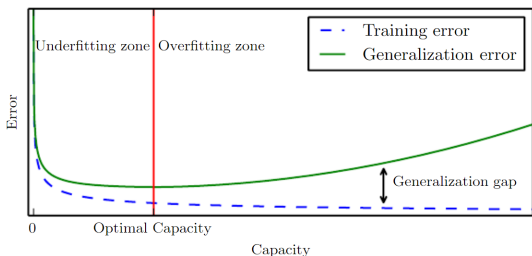
# CAPACITY



Credit: Ian Goodfellow

- The performance of a learner depends on its ability to:
  - Minimize the training error
  - Generalize well to new data
- Failure to obtain a sufficiently low training error is known as **underfitting**.
- On the other hand, if there is a large difference in training and test error, this is known as **overfitting**.

# CAPACITY



Credit: Ian Goodfellow

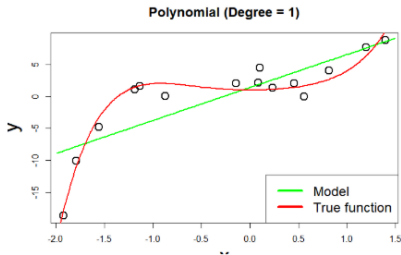
- The tendency of a model to over-/underfit is a function of its **capacity**, determined by the type of hypotheses it can learn.
- Loosely speaking, a model with low capacity can only learn a few simple hypotheses, whereas a model with large capacity can learn many, possibly complex, hypotheses.
- As the figure shows, the test error is minimized when the model neither underfits nor overfits, that is, when it has the right capacity.

# OVERFITTING

- The capacity (or “complexity”) of a model can be increased by increasing the size of the hypothesis space.
- This (usually) also increases the number of learnable parameters.
- Examples: Increasing the degree of the polynomial in linear regression, increasing the depth of a decision tree or a neural network, adding additional predictors, etc.
- As the size of the hypothesis space increases, the tendency of a model to overfit also increases.
- Such a model might fit even the random quirks in the training data, thereby failing to generalize.

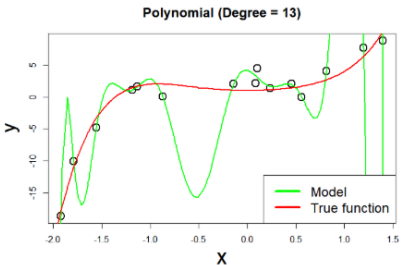
# OVERFITTING: POLYNOMIAL REGRESSION

**Degree = 1**  
(highest  
degree of a  
term in the  
polynomial)



Underfitting  
(Low Capacity)

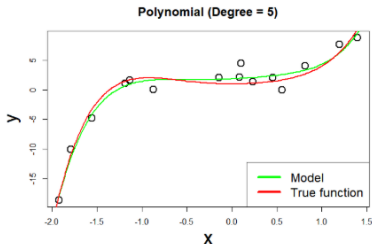
**Degree = 13**



Overfitting  
(High Capacity)

# OVERFITTING: POLYNOMIAL REGRESSION

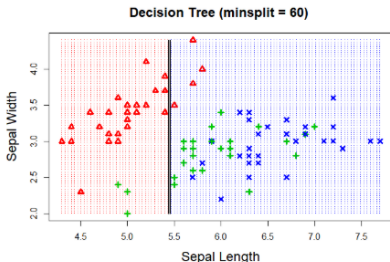
Degree = 5



	Degree = 1	Degree = 5	Degree = 13
Training error (RMSE)	3.87	1.23	0.48
Test error (RMSE)	4.11	1.55	148.5

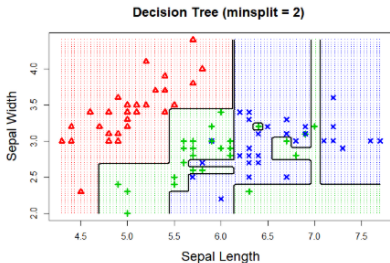
# OVERFITTING: DECISION TREES

**minsplit = 60**  
(minimum  
number of  
samples in a  
node being  
split)



Underfitting  
(Low Capacity)

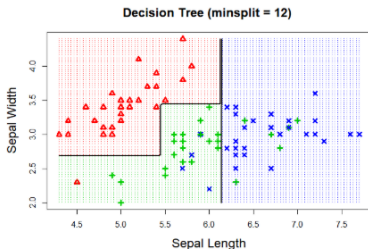
**minsplit = 2**



Overfitting  
(High Capacity)

# OVERFITTING: DECISION TREES

minsplit = 12



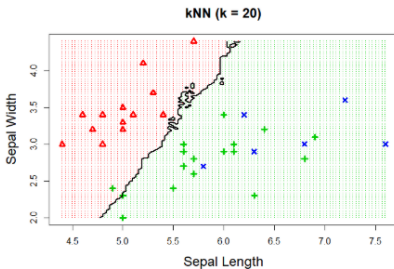
Good fit  
(Appropriate  
capacity)

	minsplit = 60	minsplit = 12	minsplit = 2
Training error (Misclassification)	0.36	0.12	0.02
Test error (Misclassification)	0.40	0.32	0.35



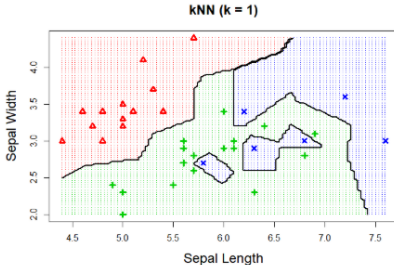
# OVERFITTING: K-NEAREST NEIGHBORS

**k = 20**  
(number of  
neighbours)



Underfitting  
(Low Capacity)

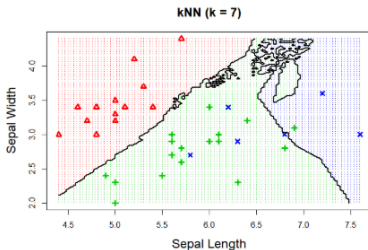
**k = 1**



Overfitting  
(High Capacity)

# OVERFITTING: K-NEAREST NEIGHBORS

$k = 7$



Good fit  
(Appropriate  
capacity)

	$k = 20$	$k = 7$	$k = 1$
Training error (Misclassification)	0.22	0.13	0
Test error (Misclassification)	0.40	0.25	0.33