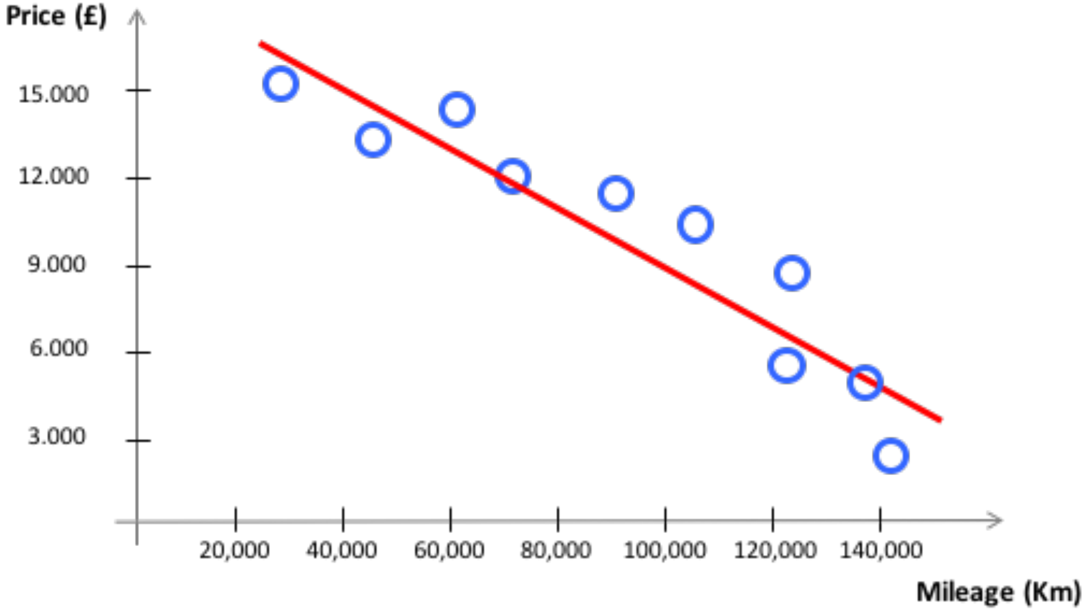
4. Supervised Learning: Polynomial Regression

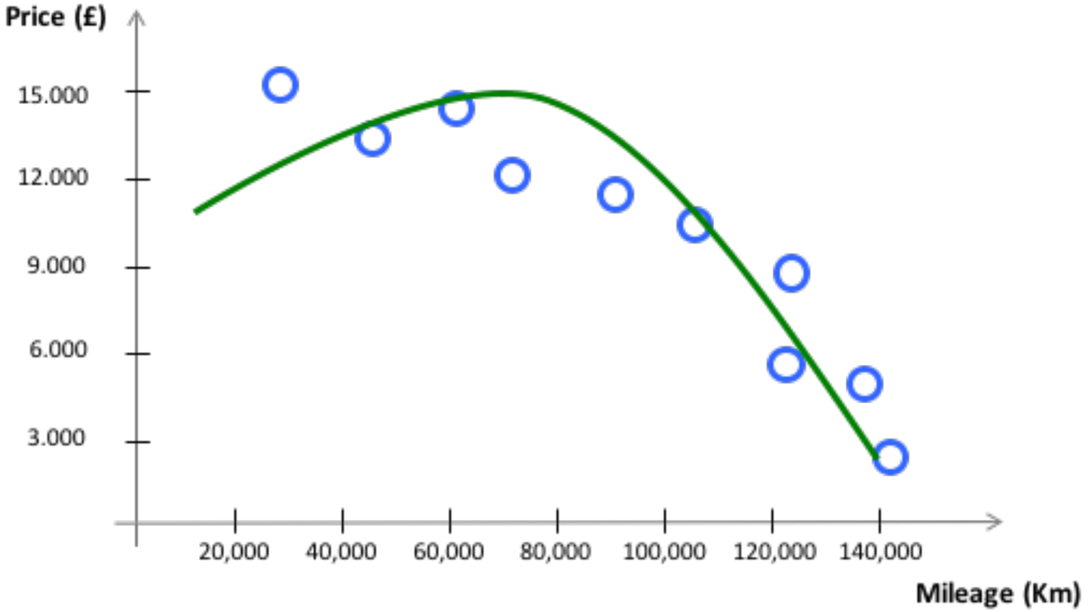
Polynomial Regression

Simple Linear Regression



First-order model: $y = \beta_0 + \beta_1 x$

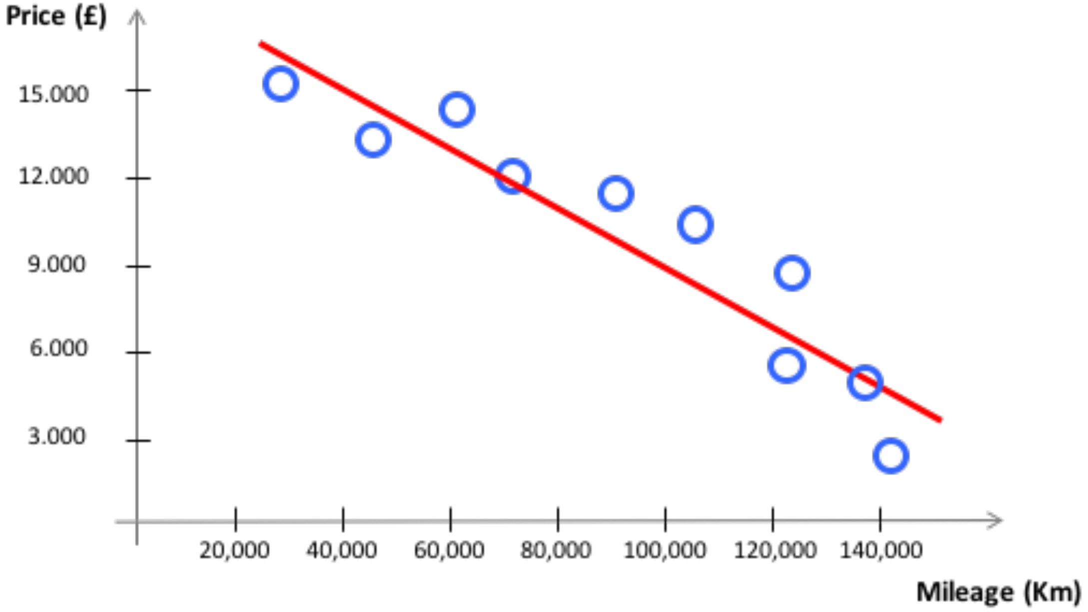
Polynomial Regression



Second order model: $y = \beta_0 + \beta_1 x + \beta_2 x^2$

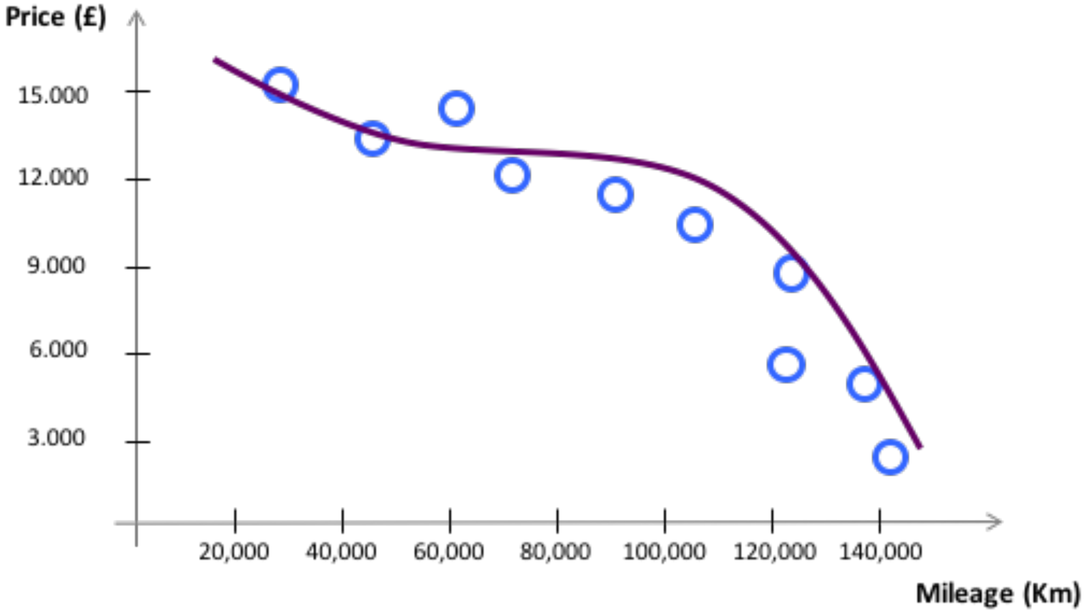
Polynomial Regression

Simple Linear Regression



First-order model: $y = \beta_0 + \beta_1 x$

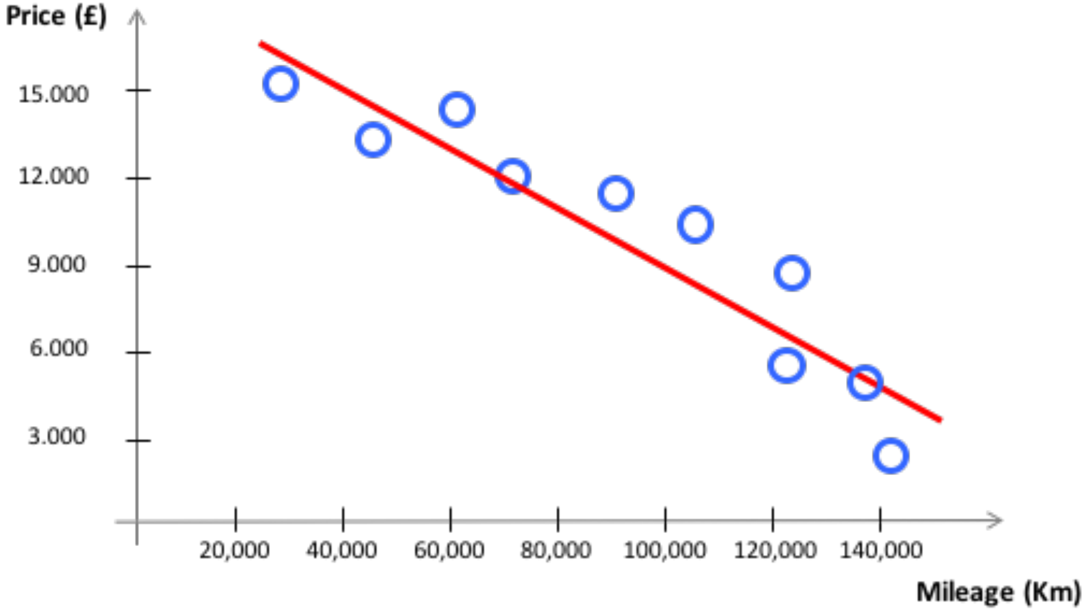
Polynomial Regression



Third order model: $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$

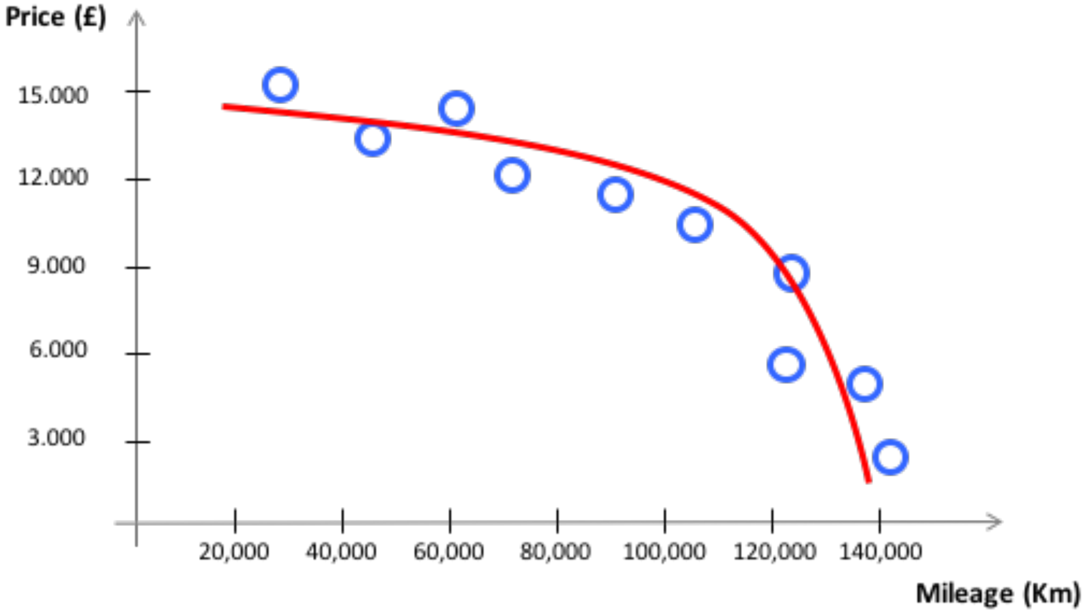
Polynomial Regression

Simple Linear Regression



First-order model: $y = \beta_0 + \beta_1 x$

Polynomial Regression



Squared-root model: $y = \beta_0 + \beta_1 \sqrt{x}$

Polynomial Regression: How to Apply?

Example: Second-order Polynomial Regression

Model

$$y = \beta_0 + \beta_1 \underbrace{x}_{x_1} + \beta_2 \underbrace{x^2}_{x_2} \Rightarrow y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \Rightarrow$$

Apply Simple Linear Regression

Feature 1 [x_1] (Mileage)	Feature 2 [x_2] (Mileage^2)	Output Variable [y] (Cost, £)
25,000	$25{,}000^2$	16,000
105,000	$105{,}000^2$	11,500
120,000	$120{,}000^2$	6,000
140,000	$140{,}000^2$	3,000
45,000	$45{,}000^2$	13,500

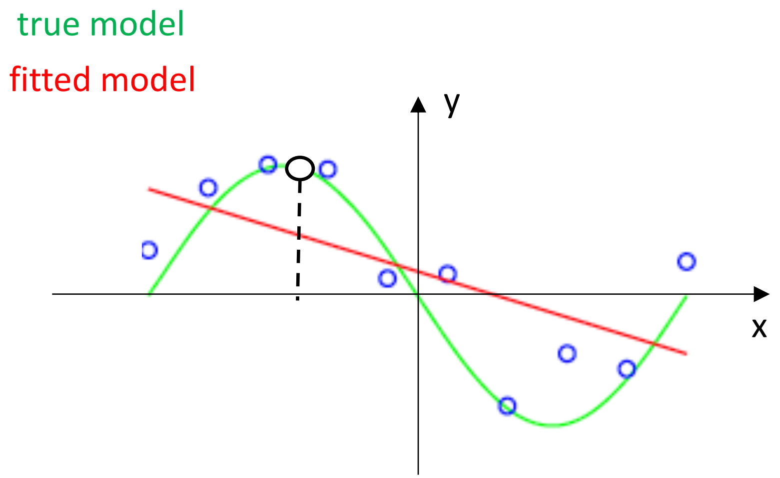
Features will have quite some different scales!
We will see later the importance in scaling them!

The Problem of Overfitting

Does a higher-order model produce better results compared to a lower-order one?

First-Order Model

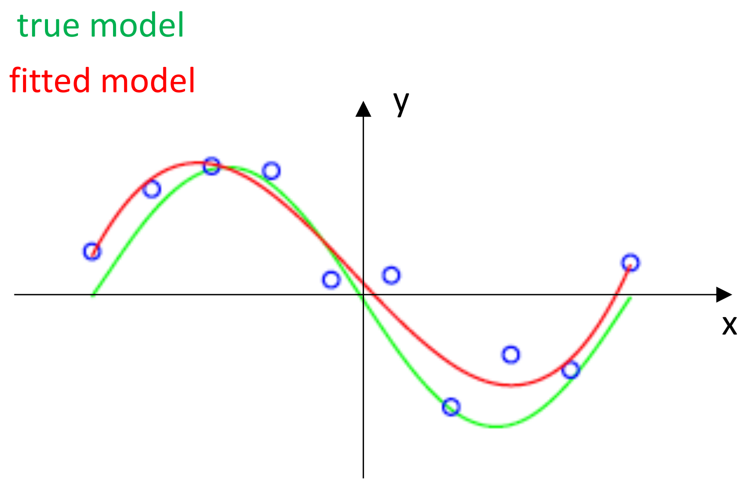
$$y = \beta_0 + \beta_1 x$$



Under-fitting model

Third-Order Model

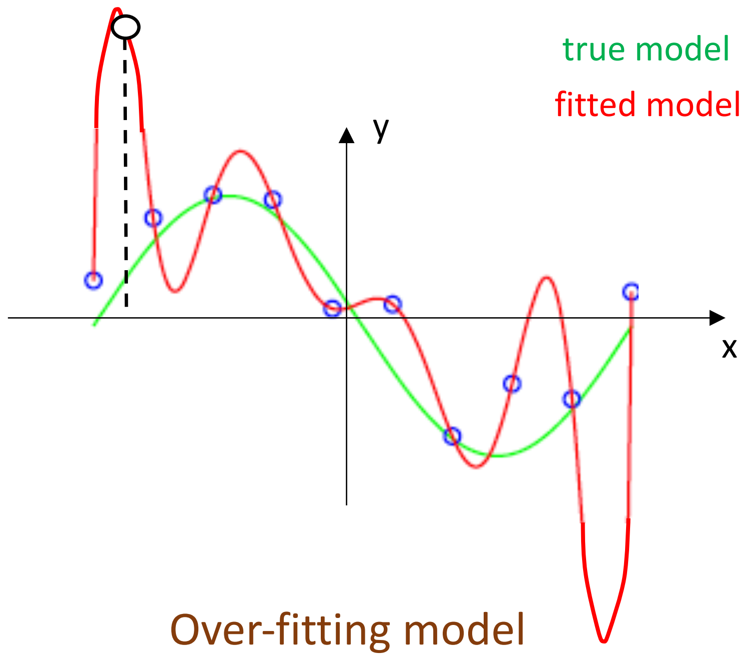
$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$$



Good fitting model

Ninth-Order Model

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_9 x^9$$



Over-fitting model

The Problem of Overfitting

The use of a small hypothesis space – such as one consisting of linear models – can lead to ***underfitting*** to the training data. It is also known as ***high-bias*** setting because we are “biasing” the learning algorithm – due to possible preconceptions about the data – to choose a very constrained model.

Conversely, the use a a large hypothesis space – such as one consisting of high-order polynomials – can lead to ***overfitting*** to the training data. It is typically known as a ***high-variance*** setting.

Both a high-bias and a high-variance setting can lead to poor prediction performance in the presence of new data.

*There is typically a **bias-variance tradeoff** in the sense that one cannot achieve both low-bias and low-variance simultaneously.*

How to Deal with Overfitting?

- Feature Selection

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cancel{\beta_3 x^3}$$

  - This can involve **manual** selection of relevant features
  - It can be problematic because such feature selection procedure can also lead to loss of valuable information for prediction purposes

- Regularization

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$$

  - This involves **automatic** reduction of the magnitude of some of the model parameters
  - It can work well in problems involving various features with each one contributing a bit to the prediction