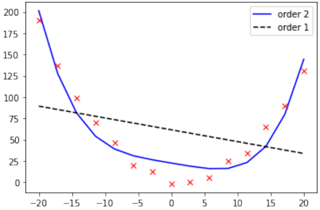
Making linear models complicated

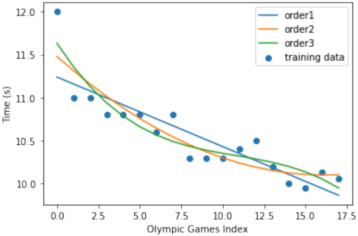
- Last week we saw how linear models could be extended to explain more complex relationships in the input data.

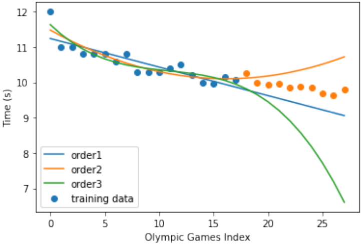
$$X = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^K \\ 1 & x_2 & x_2^2 & \dots & x_2^K \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_N & x_N^2 & \dots & x_N^K \end{bmatrix}$$

So we can fit polynomials, or any set of basis functions:



Polynomial regression



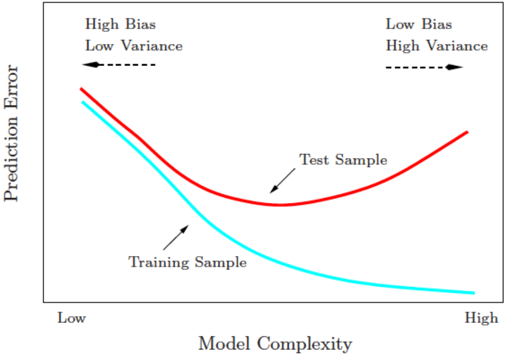


Sometimes the simplest model will be best for predicting new data.

University of Sussex

Training and Test Error as a Function of Model Complexity

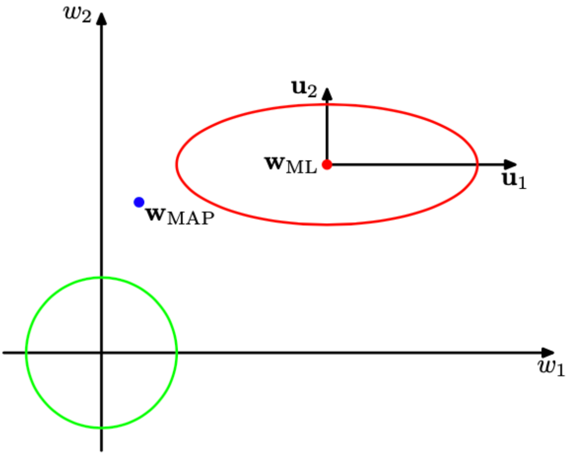
Model complexity: for e.g. degree of polynomials of the basis functions. The higher degree, the more complex.



The Elements of Statistical Learning

University of Sussex

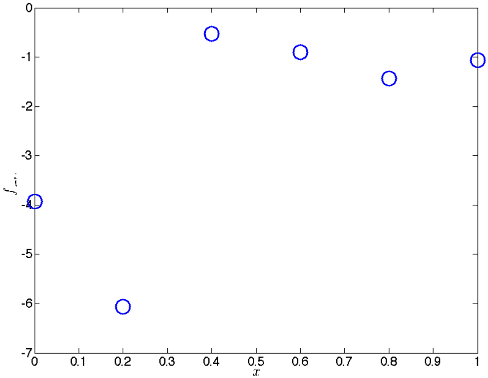
MAP vs ML



A geometric perspective. w_{map} is between w_{ml} and the prior.
Taken from Bishop: Pattern recongition and machine learning

Regularisation in Action

Our data with just 6 data points: $(0, -4), (0.2, -6), \ldots$

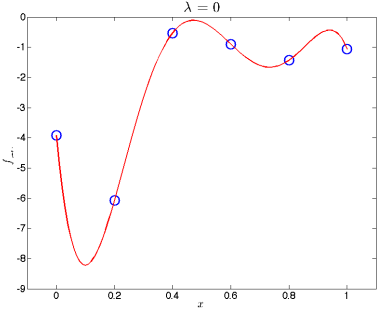


Regularisation in Action

- We use the 5th order polynomial model

$$\hat{f}(x; \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4 + w_5 x^5$$

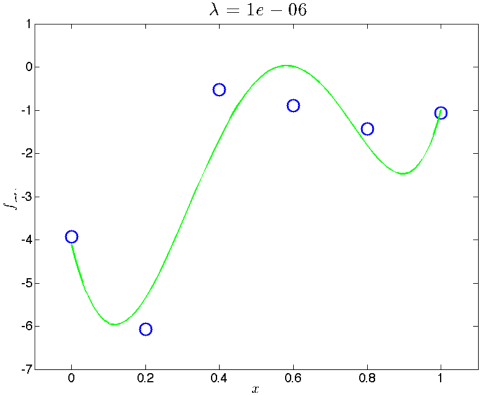
- We set $\lambda = 0 \rightarrow$ we recover the non-regularised version of linear regression



$$\underset{\mathbf{w}}{\text{minimise}} \quad \left\{ \mathcal{L}(y, \hat{f}(x; \mathbf{w})) + \frac{\lambda}{2} \mathbf{w}^\top \mathbf{w} \right\}$$

Regularisation in Action

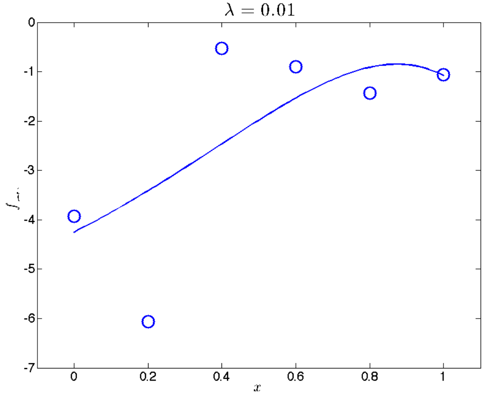
- We set $\lambda = 1e-06 \rightarrow$ the model follows the general shape of the exact 5th order polynomial but without as much variability and is further away from the data points.



$$\underset{\mathbf{w}}{\text{minimise}} \quad \left\{ \mathcal{L}(y, \hat{f}(x; \mathbf{w})) + \frac{\lambda}{2} \mathbf{w}^\top \mathbf{w} \right\}$$

Regularisation in Action

- We set $\lambda = 0.01 \rightarrow$ the model becomes less complex.

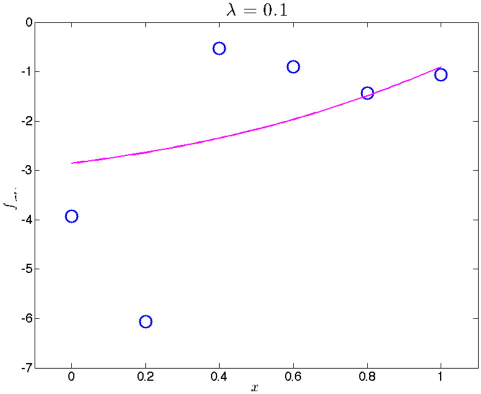


$$\underset{\mathbf{w}}{\text{minimise}} \quad \left\{ \mathcal{L}(y, \hat{f}(x; \mathbf{w})) + \frac{\lambda}{2} \mathbf{w}^\top \mathbf{w} \right\}$$

University of Sussex

Regularisation in Action

- We set $\lambda = 0.1 \rightarrow$ the model becomes even less complex.



$$\underset{\mathbf{w}}{\text{minimise}} \quad \left\{ \mathcal{L}(y, \hat{f}(x; \mathbf{w})) + \frac{\lambda}{2} \mathbf{w}^\top \mathbf{w} \right\}$$

University of Sussex