Workshop 3

COMP90051 Statistical Machine Learning

Semester 1, 2023

Learning outcomes

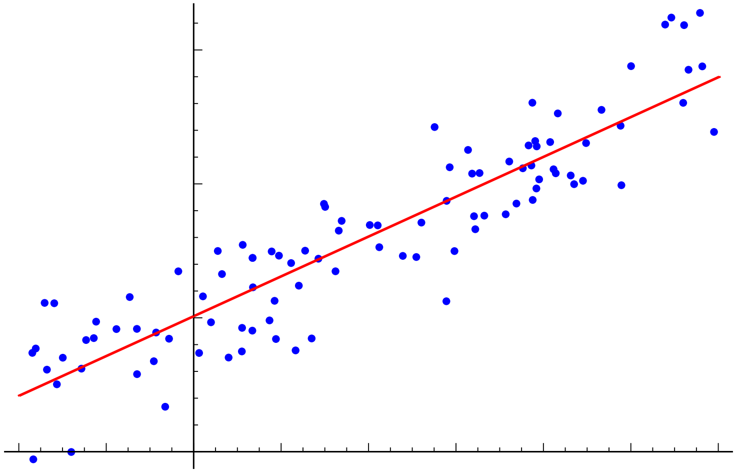
At the end of this workshop you should:

- be able to implement linear regression using numerical linear algebra functions
- be familiar with the scikit-learn interface for linear regression
- be able to implement polynomial regression
- be able to explain the benefits/drawbacks of linear regression versus polynomial regression

Quick review: linear regression

Assume the response y is a *linear* function of the features $\mathbf{x} = [x_1, \dots, x_m]^\mathrm{T}$:

$$y = w_0 + \sum_{i=1}^{m} w_i \cdot x_i$$



Write this more compactly as $y = \mathbf{x}^\mathrm{T}\mathbf{w}$ by redefining $\mathbf{x} = [x_0, x_1, \dots, x_m]^\mathrm{T}$ with $x_0 = 1$ and defining $\mathbf{w} = [w_0, \dots, w_m]^\mathrm{T}$

Question: How do we choose the weights?

Quick review: linear regression

Decision theoretic view

Make decision that minimises the empirical risk

$$\hat{R} = \frac{1}{n}\sum_{i=1}^{n} L(y_i, \hat{y}_i)$$

and choose the square loss $L(y, \hat{y}) = (\hat{y} - y)^2$.

Optimal decision for $\mathbf{w}$ minimises the sum-squared error.

Probabilistic view

Assume

$$y|\mathbf{x}, \mathbf{w} \sim \mathcal{N}(\mathbf{x}^{\mathrm{T}}\mathbf{w}; \sigma^2)$$

Can write down the likelihood for the observations

$$L(w|\boldsymbol{X}, \boldsymbol{Y}) = \prod_{i=1}^{n} p(y_i|\mathbf{x}_i, \mathbf{w}, \sigma)$$

MLE for $\mathbf{w}$ minimises the sum-squared error.

Worksheet 3