COMP90051
Statistical Machine Learning

Workshop Week 3

Xudong Han

https://github.com/HanXudong/COMP90051_2020_S1

About Me

- Xudong Han
- xudong.han@unimelb.edu.au
- Slides https://github.com/HanXudong/COMP90051_2020_S1

Review

- Bayesian versus Frequentist
- Probability density function (PDF)
- Bernoulli distribution
- Likelihood
- Log trick and log-likelihood
- MLE

Learning Outcomes

At the end of this workshop you should be able to:

- Know the assumption of linear regression.
- Derive linear regression analytic solution by hand.
- Apply nonlinear basis functions.
- Have a preview of overfitting.

Key Points

- Probability density function (pdf)
- Maximum Likelihood Estimation (MLE)
- Risk function
- Normal distribution
- Multivariate normal distribution
- Closed form solution (aka analytic solution)
- Basis functions
- Overfitting

Notations

- Random Variable

a, b, c

- Column Vector

$\vec{x}, \vec{y}, \vec{z}$

- Matrix

X, Y, Z

- Transpose

$\vec{x}^T, X^T$

Review Linear Regression

- $y_i = \overrightarrow{x_i}^T \overrightarrow{w} + \varepsilon_i$
  - $y_i \sim N(\overrightarrow{x_i}^T \overrightarrow{w}, \sigma^2)$
  - $\overrightarrow{x_i}^T = [1, x_{i1}, \dots x_{ip}]$
  - $\overrightarrow{w}^T = [w_0, w_1, \dots, w_p]$
  - $\varepsilon_i \sim N(0, \sigma^2)$

In matrix form

- Suppose we have n pairs of instances.
- $\vec{y} = [y_1, \dots, y_n]^T$
- $\overrightarrow{x_i} = [1, x_{i1}, \dots, x_{ip}]$
- $X = [\overrightarrow{x_1}, \dots, \overrightarrow{x_n}]^T$
- $\vec{w} = [w_0, \dots, w_p]^{\mathrm{T}}$
- By the definition of linear regression, we have the following equation:

$$\vec{y} = X\vec{w} + \vec{\varepsilon}$$

where $\vec{\varepsilon} \sim Multivariate\ Normal(\vec{0}, \sigma^2 I_{n \times n})$.

Analytic Solution

- MLE: Maximise Likelihood

$$\widehat{\vec{w}} = argmax \prod_k \boldsymbol{\mathcal{N}}(y_k; \overrightarrow{x_k}^T w, \sigma^2)$$

$$\widehat{\vec{w}} = argmax\, \boldsymbol{\mathcal{N}}(\vec{y}; X\vec{w}, \sigma^2 I_{n\times n})$$

- Decision Theory: Minimise risk function

$$\widehat{\vec{w}} = argmin \sum_k \left(y_k - \overrightarrow{x_k}^T w\right)^2$$

$$\widehat{\vec{w}} = argmin\, [(\vec{y} - X\vec{w})^{\mathrm{T}}(\vec{y} - X\vec{w})]$$

Solve for the optimal weights $\mathbf{w}^\star$ analytically

- MLE: Maximise Likelihood

$$\widehat{\vec{w}} = argmax\, \boldsymbol{\mathcal{N}}(\vec{y}; X\vec{w}, \sigma^2 I_{n\times n})$$

- Decision Theory: Minimise risk function

$$\widehat{\vec{w}} = argmin\, [(\vec{y} - X\vec{w})^{\mathrm{T}}(\vec{y} - X\vec{w})]$$

Nonlinear Basis Function

- Specifically, we are using polynomial expansions here.
- For more details, https://web.stanford.edu/~hastie/Papers/ESLII.pdf
- Pros and Cons

Overfitting