Short course

A vademecum of statistical pattern recognition and machine learning

Linear regression: a brief introduction

Massimo Piccardi
University of Technology, Sydney, Australia

Agenda

- Linear regression
- Minimum least squares
- Gaussian regression, MLE
- Multiple regression
- Multivariate regression

Linear regression

- Assume that we have two random variables y, x $\in \Re$: a **linear regressor** from x (*input*, *independent variable, explanatory variable etc*) to y (*output*, *dependent variable etc*) is a linear function attempting to predict a value for y from a value for x:

$$\tilde{y} = w\,x + b$$

- The above is called *simple linear regression*, with w and b scalar parameters, and b sometimes called *bias*

Linear regression

- We call *error*, ε, the difference between y and its prediction, ỹ:

$$\varepsilon = y - \tilde{y} = y - (w\,x + b)$$

$$y = w\,x + b + \varepsilon$$

Compacted notation

- The previous notation has the bias, b, explicit. We can compact the notation by considering:

$$w' = \begin{bmatrix} w \\ b \end{bmatrix}, \; x' = \begin{bmatrix} x \\ 1 \end{bmatrix}$$

$$\rightarrow \tilde{y} = w'^{T} \; x'$$

- In these new co-ordinates, the bias term is included in the weight vector

Least squares estimation

- Assume that we have a set of pairs $\{x_i, y_i\}$, i = 1...N, and that we want to estimate the best w,b from this training set
- As objective function, we choose the total square error between the ground-truth value, y_i, and its prediction, $\tilde{y}_i$:

$$\sum_{i=1}^{N} (y_i - \tilde{y}_i)^2 = \sum_{i=1}^{N} (y_i - (w\,x_i + b))^2$$

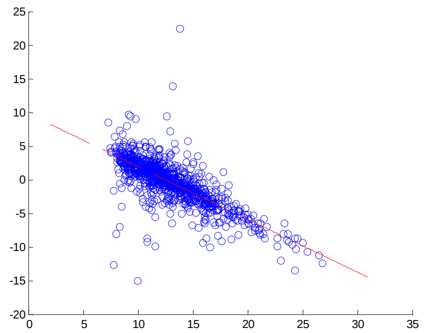
Least squares estimation

- We note as X and Y two N x 1 vectors concatenating all inputs and outputs, respectively (*data vectors*)
- We also compact the notation as shown previously. X is now an N x 2 vector, w is 2 x 1
- The least-square solution is:

$$w_{LS} = \left(X^T X\right)^{-1} X^T Y$$

Example

- $x_i \sim \text{Gamma}(2,3) + 7$ $y_i \sim t(\mathbf{-0.8} * x_i + \mathbf{10},2,2)$



- Least-square estimate: w = -0.7923; b = 9.9315

Multiple linear regression

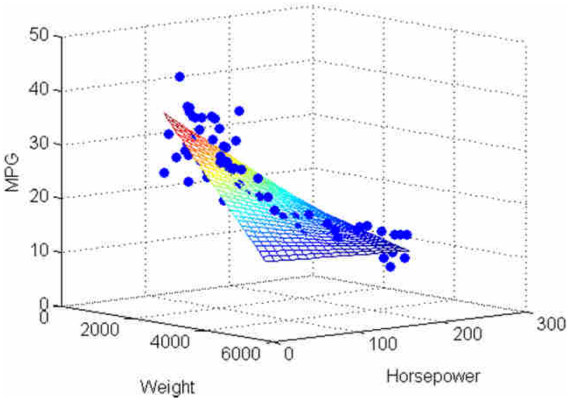
- Variable y can be obtained as the linear regression of a multivariate input, $x \in \Re^P$:

$$\tilde{y} = w_1 \; x_1 + \ldots + w_P \; x_P + b =$$
$$= w^T x + b$$

- We can again extend $w = [w_1 \ldots w_P]^T$ with b and $x = [x_1 \ldots x_P]^T$ with 1: **the least-square solution is unchanged**

- This case is called *multiple linear regression*

Example



- Predicts car's mileage per gallon from weight and horsepower (source: the MathWorks)

Least-square solution

- We can also re-write the least-square solution without using data matrices X and Y:

$$w_{LS} = \left(\sum_{i=1}^{N} x_i x_i^T \right)^{-1} \sum_{i=1}^{N} x_i y$$

- Just as a property, note that if w is transposed we have:

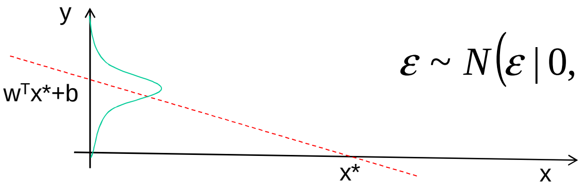
$$w_{LS}^T = \left(\left(X^T X \right)^{-1} X^T Y \right)^T = \left(X^T Y \right)^T \left(X^T X \right)^{-1^T} =$$

$$= Y^T X \left(X^T X \right)^{-1} = \sum_{i=1}^{N} y x_i^T \left(\sum_{i=1}^{N} x_i x_i^T \right)^{-1}$$

Gaussian assumption

- We now extend our assumptions over y by assuming that it is Gaussian-distributed conditioned on x:

$$y = w^T x + b$$
$$\rightarrow y \sim N\left(y \mid w^T x + b, \sigma^2 \right)$$



$$\varepsilon \sim N\left(\varepsilon \mid 0, \sigma^2 \right)$$

MLE

- The maximum-likelihood estimate (MLE) for the aggregated [w,b] is identical to that of least squares!
- We also add an estimate for the variance around the prediction, σ^2

$$w_{ML} = \left(\sum_{i=1}^{N} x_i x_i^T \right)^{-1} \sum_{i=1}^{N} x_i y$$

$$\sigma_{ML}^2 = \frac{1}{N} \sum_{i=1}^{N} \left(y_i - w_{ML}^T x_i \right)^2$$

Multivariate linear regression

- Now assume $y \in \Re^Q$: the case is called **multivariate linear regression**
- The conditional Gaussian model can be written as

$$y \sim N\left(y \mid W^T x + b, \Sigma \right)$$

W is a P x Q matrix, b is a Q x 1 vector and Σ is a Q x Q covariance matrix

MLE

- MLE solution:

$$W_{ML} = \left(\sum_{i=1}^{N} x_i x_i^T \right)^{-1} \sum_{i=1}^{N} x_i y_i^T$$

$$\Sigma_{ML} = \frac{1}{N} \sum_{i=1}^{N} \left(y_i - W_{ML}^T x_i \right) \left(y_i - W_{ML}^T x_i \right)^T$$