Derivation of closed form solution

September 30, 2018

1 Introduction

Linear regression function $y(x, w)$ has the form:

$$y(x, w) = w^T \phi(x) \tag{1}$$

where $w = (w_0, w_1, .., w_{M1})$ is a weight vector to be learnt from training samples and $\phi = (\phi_0, .., \phi_{M-1}^T)$ is a vector of M basis functions.
Assuming $\phi_0(x) = 1$ for whatever input, w_0 becomes the bias term. Each basis function $\phi(x)$ converts the input vector x into a scalar value.

Objective function

Minimizing the sum-of-squares error:

$$E_D(w) = \frac{1}{2} \sum_{n=1}^{N} (t_n - w^T \phi(x_n))^2 \tag{2}$$

where $t = t_1, .., t_N$ is the vector of outputs in the training data and Φ is the design matrix.

2 Overview of differentiation

$$\sum e^2 = e'e \tag{3}$$

where e is a vector.

$$\frac{\partial a'b}{\partial b} = \frac{\partial b'a}{\partial b} = a \tag{4}$$

when a and b are Kx1 vectors.

$$\frac{\partial b'Ab}{\partial b} = 2Ab = 2b'A \tag{5}$$

when A is any symmetric matrix.

$$\frac{\partial 2\beta' X' y}{\partial b} = \frac{\partial 2\beta'(X'y)}{\partial b} = 2X'y \tag{6}$$

$$\frac{\partial \beta' X' X\beta}{\partial b} = \frac{\partial \beta' A\beta}{\partial b} = 2A\beta = 2X'X\beta \tag{7}$$

when $X'X$ is a ExK matrix.

$$y'X\beta = (y'X\beta)' = \beta' X' y \tag{8}$$

3 Derivation of closed form solution

Sum of squares error:

$$\begin{aligned} E(w) &= \frac{1}{2}\sum_{n=1}^{N}(t_n - w^T\Phi_n)^2 \\ &= \frac{1}{2}(t - \Phi w)'(t - \Phi w) \\ &= \frac{1}{2}(t't - w'\Phi' t - t'\Phi w + w'\Phi'\Phi w) \\ &= \frac{1}{2}(t't - 2w'\Phi' t + w'\Phi'\Phi w) \end{aligned} \tag{9}$$

To find w that minimize the sum of squared error, we need to take the derivative of equation above with respect to w. This gives us the following:

$$\frac{\partial E(w)}{\partial w} = -\Phi' t + \Phi'\Phi w \tag{10}$$

From previous equation we get "normal equations":

$$\Phi'\Phi w = \Phi' t \tag{11}$$

Recall, that $\Phi'\Phi$ is a square symmetric matrix kxk. $\Phi'\Phi$ and $\Phi' t$ is known, but w is unknown. If the inverse of $\Phi'\Phi$ exists, then pre-multiplying both sides by this inverse we get:

$$(\Phi'\Phi)^{-1}(\Phi'\Phi)w = (\Phi'\Phi)^{-1}\Phi' t \tag{12}$$

By definition, $(\Phi'\Phi)^{-1}(\Phi'\Phi) = I$, where I is kxk identity matrix. Thus

$$Iw = (\Phi'\Phi)^{-1}\Phi' t \tag{13}$$

$$w = (\Phi'\Phi)^{-1}\Phi' t \tag{14}$$