EE2100: Matrix Analysis

Review Notes - 25

Topics covered :

1. Linear regression

---

1. **Linear Regression**: Let $\left[\left(y^1, x_1^1, x_2^1, \cdots, x_n^1\right), \left(y^2, x_1^2, x_2^2, \cdots, x_n^2\right), \cdots, \left(y^m, x_1^m, x_2^m, \cdots, x_n^m\right)\right]$ denote the training data set. The input in this scenario has n features and can be denoted by a vector $\mathbf{x} = [x_1, x_2, \cdots, x_n]^T$. The output/target which we want to predict is denoted by y. The training data point $\left(y^i, x_1^i, x_2^i, x_3^i, \cdots, x_n^i\right)$ indicates the data corresponding to measurement i (the superscript is an indication that the data corresponds to i^{th} measurement). The objective of linear regression is to compute a polynomial (say $h(\mathbf{x})$; often referred to as hypothesis function), given by

$$\begin{aligned} y = h(\mathbf{x}) &= c_0 + \mathbf{c}^T\mathbf{x} \\ &= c_0 + c_1 x_1 + \cdots + c_n x_n \end{aligned} \tag{1}$$

In order to determine the linear hypothesis function that best fits the given data, it is necessary to determine the coefficients $c_0, \cdots, c_n$. The chosen hypothesis, when evaluated at the inputs corresponding to the training data results in

$$\begin{aligned} c_0 + c_1 x_1^1 + c_2 x_2^1 + \cdots + c_n x_n^1 &= y^1 \\ \vdots \qquad\qquad &\quad \vdots \\ c_0 + c_1 x_1^m + c_2 x_2^m + \cdots + c_n x_n^m &= y^1 \end{aligned} \tag{2}$$

The set of equations given by (2) can be represented as

$$\underbrace{\begin{bmatrix} 1 & x_1^1 & x_2^1 & \cdots & x_n^1 \\ 1 & x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_1^m & x_2^2 & \cdots & x_n^m \end{bmatrix}}_{\mathbf{A} \in \mathcal{R}^{m \times n+1}} \underbrace{\begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{bmatrix}}_{\mathbf{c} \in \mathcal{R}^{n+1}} = \underbrace{\begin{bmatrix} y^1 \\ y^2 \\ \vdots \\ y^m \end{bmatrix}}_{\mathbf{y} \in \mathcal{R}^m} \tag{3}$$

Equation (3) represents an overdetermined set of linear equations (if the number of data points are greater than the order of polynomial) and hence an optimal value of coefficients of the hypothesis function can be computed by solving the linear equation

$$\left(\mathbf{A}^t\mathbf{A}\right)\mathbf{c} = \mathbf{A}^t\mathbf{y} \tag{4}$$