Lecture 13: Basis Function Regression

Course Teacher: Md. Shariful Islam Bhuyan

Regression

- Multivariate linear regression

$$f(\mathbf{x}) = w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_d x_d = \mathbf{w}^{\mathrm{T}}\mathbf{x} \qquad \mathbf{w} \in \mathbb{R}^{\mathrm{d}}$$

$$L\big(y, f(\mathbf{x})\big) = \left(y - \mathbf{w}^{\mathrm{T}}\mathbf{x}\right)^2$$

$$\mathbf{w}^* = \arg\min_{f} \mathcal{L}_{emp}(f) = \arg\min_{\mathbf{w}\in\mathbb{R}^{\mathrm{d}}} \sum_{i=1}^{N} \left(y_i - \mathbf{w}^{\mathrm{T}}\mathbf{x}_i\right)^2$$

$$= \arg\min_{\mathbf{w}\in\mathbb{R}^{\mathrm{d}}} \left\| \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} - \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{d1} \\ 1 & x_{12} & x_{22} & \cdots & x_{d2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{1N} & x_{2N} & \cdots & x_{dN} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix} \right\|^2 = \arg\min_{\mathbf{w}\in\mathbb{R}^{\mathrm{d}}} \|\mathbf{y} - \mathbf{X}\mathbf{w}\|^2$$

- Extension: Multivariate polynomial regression, same objective function

Adaptive Basis Function

- What about this? Still linear! Feature transformation.

$$f(\mathbf{x}) = w_0 + w_1 x_1 + w_2 x_2^3 + w_3 \sqrt{x_3} + w_4 e^{-x_4} + w_5 \ln x_5 + \cdots + w_d \sin^{-1} x_d = \mathbf{w}^{\mathrm{T}} \mathbf{x}$$

- Not linear!

$$f(\mathbf{x}) = w_0 + w_1 x_1 x_2$$

- Adaptive Basis Function Model (also contains decision tree, very general model)

$$f(\mathbf{x}) = \sum_{k=0}^{d} w_k \phi_k(\mathbf{x})$$

Analytical Solution

$$\frac{\partial \mathcal{L}_{emp}}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \|\mathbf{y} - \mathbf{X}\mathbf{w}\|^2 = \frac{\partial}{\partial \mathbf{w}} (\mathbf{y} - \mathbf{X}\mathbf{w})^{\mathrm{T}} (\mathbf{y} - \mathbf{X}\mathbf{w}) = \frac{\partial}{\partial \mathbf{w}} \left(\mathbf{y}^{\mathrm{T}} - \mathbf{w}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\right)(\mathbf{y} - \mathbf{X}\mathbf{w})$$

$$= \frac{\partial}{\partial \mathbf{w}} \left(\mathbf{y}^{\mathrm{T}}\mathbf{y} - \mathbf{w}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{y} - \mathbf{y}^{\mathrm{T}}\mathbf{X}\mathbf{w} + \mathbf{w}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{w}\right) \quad = \quad \frac{\partial}{\partial \mathbf{w}} \left(\mathbf{y}^{\mathrm{T}}\mathbf{y} - \mathbf{w}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{y} - \left(\mathbf{w}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{y}\right)^{\mathrm{T}} + \mathbf{w}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{w}\right)$$

$$= \frac{\partial}{\partial \mathbf{w}} \left(\mathbf{y}^{\mathrm{T}}\mathbf{y} - \mathbf{2}\mathbf{w}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{y} + \mathbf{w}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{w}\right) = \frac{\partial}{\partial \mathbf{w}} \mathbf{y}^{\mathrm{T}}\mathbf{y} - \frac{\partial}{\partial \mathbf{w}} \mathbf{2}\mathbf{w}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{y} + \frac{\partial}{\partial \mathbf{w}} \mathbf{w}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}$$

$$\frac{\partial}{\partial \mathbf{w}} \mathbf{w}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{w} = 2\mathbf{w}^{\mathbf{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X} = 2\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{w} \qquad \left\{\frac{\partial}{\partial \mathbf{x}} \mathbf{x}^{T}\mathbf{A}\mathbf{x} = \mathbf{x}^{T}(\mathbf{A} + \mathbf{A}^{T})\right\}$$

$$\frac{\partial}{\partial \mathbf{w}} \mathbf{w}^{\mathrm{T}} = \mathbf{I}_{d+1}$$

$$\frac{\partial \mathcal{L}_{emp}}{\partial \mathbf{w}} = -\mathbf{2}\mathbf{X}^{\mathrm{T}}\mathbf{y} + 2\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{w} = 0, \qquad \mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{w} - \mathbf{X}^{\mathrm{T}}\mathbf{y} = 0, \qquad \mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{w} = \mathbf{X}^{\mathrm{T}}\mathbf{y}, \qquad \mathbf{w} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-\mathbf{1}}\mathbf{X}^{\mathrm{T}}\mathbf{y}$$