3/17/2025
Monday, March 17, 2025 11:59 AM

Nonlinear Regression

- Consider

$$\hat{\theta} = \underset{\theta \in \Gamma}{\arg\min} \; J(\theta) \equiv \sum_{i=1}^{N_y} \left(y_i - \underline{f}(x_i ; \vec{\theta}) \right)^2$$

given by OLS solution, i.e. $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

Estimator Mean

- using gradients, we have

$$\nabla_\theta J(\theta) = -2 \nabla_\theta f(x ; \vec{\theta}) \left(\vec{y} - f(x ; \vec{\theta}) \right)$$

$$= -2 \underset{\sim}{S} \left(\vec{y} - f(\vec{x} ; \vec{\theta}) \right), \quad \underset{\sim}{S} = \frac{\partial f}{\partial \theta}$$

- If $\vec{\theta} \approx \vec{\theta}^0$, then

$$f(x ; \vec{\theta}) \approx f(x ; \vec{\theta}^0) + \left.\frac{\partial f}{\partial \vec{\theta}}\right|_{\vec{\theta}^0} \left(\vec{\theta} - \vec{\theta}^0 \right)$$

$$\Rightarrow \nabla_\theta J(\vec{\theta}) = \underset{=}{\nabla_\theta} \sum \left(y - f(x ; \vec{\theta}^0) - \underset{\sim}{S}\Big|_{\vec{\theta}^0} \cdot \left(\vec{\theta} - \vec{\theta}^0 \right) \right)^2$$

$$\vec{y} = f(x ; \vec{\theta}^0) + \vec{\varepsilon}$$

$$\Rightarrow \nabla_\theta J(\hat{\theta}) \approx -2 \underset{\sim}{S} \left(y - f(x ; \vec{\theta}^0) - \underset{\sim}{S}\Big|_{\vec{\theta}^0} \left(\hat{\vec{\theta}} - \vec{\theta}^0 \right) \right)$$

$(y = f + \vec{\varepsilon})$

$$\Rightarrow \nabla_\theta J(\hat{\theta}) \approx -2 \underset{\sim}{S} \left(\vec{\varepsilon} - \underset{\sim}{S}\Big|_{\vec{\theta}^0} \left(\hat{\vec{\theta}} - \vec{\theta}^0 \right) \right)$$

$$\Rightarrow \left(\underset{\sim}{S}^T \underset{\sim}{S} \right)^{-1} \underset{\sim}{S}^T \vec{\varepsilon} = \vec{\theta} - \vec{\theta}^0$$

- If number of observations, $N_y \to \infty$, then estimator $\hat{\vec{\theta}}$, gives

$$\hat{\vec{\theta}} - \vec{\theta}^{\circ} \approx (\underset{\sim}{S}^T \underset{\sim}{S})^{-1} \underset{\sim}{S}^T \vec{\varepsilon}, \qquad \underset{\sim}{S} = \left.\frac{\partial f}{\partial \vec{\theta}}\right|_{\vec{\theta}^{\circ}}$$

- <u>Note:</u>

$$E[\hat{\vec{\theta}}] = E[\vec{\theta}^{\circ}] + E\left[(\underset{\sim}{S}^T \underset{\sim}{S})^{-1} \underset{\sim}{S}^T \vec{\varepsilon}\right]$$

$$= E[\vec{\theta}^{\circ}] + (\underset{\sim}{S}^T \underset{\sim}{S})^{-1} \underset{\sim}{S}^T \xcancel{E[\vec{\varepsilon}]} \rightarrow 0$$

$$= E[\vec{\theta}^{\circ}]$$

$$Var[\hat{\vec{\theta}}] = E\left[(\hat{\vec{\theta}} - \vec{\theta}^{\circ})(\hat{\vec{\theta}} - \vec{\theta}^{\circ})^T\right] \approx \sigma_0^2 (\underset{\sim}{S}^T \underset{\sim}{S})^{-1}$$

where $\sigma_0^2 \approx \dfrac{1}{N_y - P} (\vec{y} - f(\underset{\sim}{x}; \hat{\vec{\theta}}))^T (\vec{y} - f(\underset{\sim}{x}; \hat{\vec{\theta}}))$

where P is the number of parameters.

- Recall for linear regression, $\hat{\vec{\theta}}_{LR} \sim N\left(\vec{\theta}^{\circ}_{LR}, \sigma_0^2 (\underset{\sim}{X}^T \underset{\sim}{X})^{-1}\right)$
- For nonlinear regression

$$\hat{\vec{\theta}}_{NR} \sim N\left(\vec{\theta}^{\circ}_{NR}, \sigma_0^2 (\underset{\sim}{S}^T \underset{\sim}{S})^{-1}\right)$$

<u>Confidence intervals</u>

- Let $v_q = (\underset{\sim}{S}^T \underset{\sim}{S})^{-1}_{qq}$

$$\Rightarrow \quad CI(\hat{\theta}_q) = \left[\hat{\vec{\theta}}_q \pm t^{1-\alpha/2}_{N_y - P} \sqrt{\sigma^2 v_q}\right]$$

$$\Rightarrow \quad CI(\hat{\theta}_q) = \left[\hat{\vec{\theta}}_q \pm t_{N_y-p}^{1-\alpha/2} \sqrt{\sigma^2 v_q}\right]$$

Uncertainty in Outputs

- Note that in linear regression, $\vec{x}^*$ represented new values of the design matrix, i.e. $\hat{y}^* = \vec{x}^* \hat{\vec{\theta}}$

- To linearize $f(\underset{\sim}{x}; \hat{\vec{\theta}})$, we replace design matrix approach with

$$\vec{g}^*\Big|_{\underset{\sim}{x}^*} = \frac{\partial f}{\partial \vec{\theta}}\Big|_{\underset{\sim}{x}^*}$$

$$\Rightarrow \quad CI(\hat{y}^*) = \left[\hat{y} \pm t_{N_y-p}^{1-\alpha/2} \sqrt{\sigma^2 \, \vec{g}^*\Big|_{\underset{\sim}{x}^*} (\underset{\sim}{S}^T \underset{\sim}{S})^{-1} \vec{g}^{*T}\Big|_{\underset{\sim}{x}^*}}\right]$$

$$\Rightarrow \quad PI(\hat{y}^*) = \left[\hat{y} \pm t_{N_y-p}^{1-\alpha/2} \sqrt{\sigma^2 + \sigma^2 \, \vec{g}^*\Big|_{\underset{\sim}{x}^*} (\underset{\sim}{S}^T \underset{\sim}{S})^{-1} \vec{g}^{*T}\Big|_{\underset{\sim}{x}}}\right]$$