Parameter CI (Linear Models), $\vec{y} = \underset{\sim}{X}\vec{\theta} + \vec{\varepsilon}$

- Once we find $\hat{\vec{\theta}} = (\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T\vec{y}$, we compute

$$\sigma^2 = \frac{1}{N_y - P}\cdot \vec{R}^T\vec{R}, \quad \vec{R} = \vec{y} - \underset{\sim}{X}\hat{\vec{\theta}}$$

$$\Rightarrow CI(\theta_q) = \left[\theta_{q,OLS} \pm t^{1-\alpha/2}_{N_y - P}\cdot \underbrace{\sqrt{\sigma^2(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}_{qq}}}_{\text{Standard error}}\right]$$

Output Uncertainty

- Let $\vec{Y} = \underset{\sim}{X}\vec{\Theta} + \vec{\varepsilon}$

- Suppose we get new input $\vec{x}_*$, & we want $\hat{y}_*$

- We can show

$$\hat{y}_* = \vec{x}_*\hat{\vec{\Theta}}$$

$$\Rightarrow Var[\hat{y}_*] = \hat{\sigma}^2\left[\vec{x}_*(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\vec{x}_*^T\right]$$

- The prediction, $\hat{y}_*$, is Normally distributed given assumptions about $\hat{\varepsilon}$.
  - We again use a T-distribution

$$T = \frac{\hat{y}_* - \mu_{x^*}}{\sqrt{\sigma^2\vec{x}_*(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\vec{x}_*^T}}, \quad \mu_{x^*}: \text{mean of model at } x^*$$

$$\Rightarrow CI(\hat{y}_{\vec{x}^*}) = \left[\underset{\uparrow \text{ will be } \vec{x}_*\hat{\vec{\Theta}}_{OLS}}{\hat{y}_*} \pm t^{1-\alpha/2}_{N_y - P}\underbrace{\sqrt{\sigma^2\vec{x}_*(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\vec{x}_*^T}}\right]$$

- Prediction intervals take into account new observations, $\hat{y}^*$, with added noise variance. (More uncertainty)

$$\Rightarrow PI(\hat{y}_{\vec{x}^*}) = \left[\hat{y}_* \pm t^{1-\alpha/2}_{N_y - P}\sqrt{\sigma^2 + \sigma^2\vec{x}_*(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\vec{x}_*}\right.$$

$$\Rightarrow PI(\hat{y}_{\vec{x}_*}) = \left[\vec{y}_* \pm t_{N_y - p} \underbrace{\sqrt{\sigma^2 + \sigma^2 \vec{x}_*(\underset{\sim}{X}^T \underset{\sim}{X})^{-1} x_*}}_{\sigma\sqrt{1 + \vec{x}_*(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\vec{x}_*^T}} \right.$$

Nonlinear Regression

- Consider

$$\hat{\Theta} = \underset{\Theta \in \Gamma}{\operatorname{argmin}} \; J(\Theta) = \sum_{i=1}^{N_y} \left(y_i - f(\underset{\sim}{x}_i ; \vec{\Theta}) \right)^2$$

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

Estimator Mean

- using gradients, we have

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

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

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

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

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

$$\Rightarrow \nabla_\theta J(\vec{\theta}) = \nabla_\theta \Big(\ldots \Big)$$