Linear model $\quad y = X\theta + \varepsilon$

$X \in \mathbb{R}^{n \times p} \quad \theta \in \mathbb{R}^{p} \quad \varepsilon \sim N(0, \Sigma)$

Likelihood $\quad L(\theta, \Sigma, y) = P(y; \theta, \Sigma)$

MLE $\quad (\hat{\theta}, \hat{\Sigma}) = \underset{\theta, \Sigma}{\operatorname{argmax}}\, L(\theta, \Sigma, y)$

assume $\Sigma = \sigma^2 I$, $\quad X \in \mathbb{R}^{n \times p}$ has full column rank

$\hat{\theta} = (X^T X)^{-1} X^T y \qquad \hat{\sigma}^2 = \frac{1}{n} \| y - X\hat{\theta} \|^2$

$\mathbb{E}[\hat{\theta}] = \theta, \quad Cov(\hat{\theta}) = \sigma^2 (X^T X)^{-1}$

Bias - Variance tradeoff

$$\mathbb{E}[\|Z\|^2] = tr(Cov(Z)) + \mathbb{E}[Z]^T\mathbb{E}[Z]$$

$$\mathbb{E}_\theta[\|\hat{\theta}-\theta\|^2] = tr(Cov(\hat{\theta}-\theta)) + \mathbb{E}[\hat{\theta}-\theta]^T\mathbb{E}[\hat{\theta}-\theta]$$

$$= \sum_{i=1}^{p} Var(\hat{\theta}_i - \theta_i) + \sum_{i=1}^{p} \mathbb{E}[\hat{\theta}_i - \theta_i]^2$$

$$= \sum_{i=1}^{p} Var(\hat{\theta}_i) + \sum_{i=1}^{p} (\mathbb{E}[\hat{\theta}_i] - \theta_i)^2$$

MSE = Variance + Bias^2

Side remark: if $Y_{new} = \theta^T x_{new} + \varepsilon_{new}$

then $\mathbb{E}[Y_{new}] = \theta^T x_{new}$ and

$\mathrm{Cov}(Y_{new}) = \sigma^2 \left(x_{new}^T (X^T X)^{-1} x_{new} + 1 \right)$

ε_{new} is independent from ε_i, $i = 1, \dots, p$