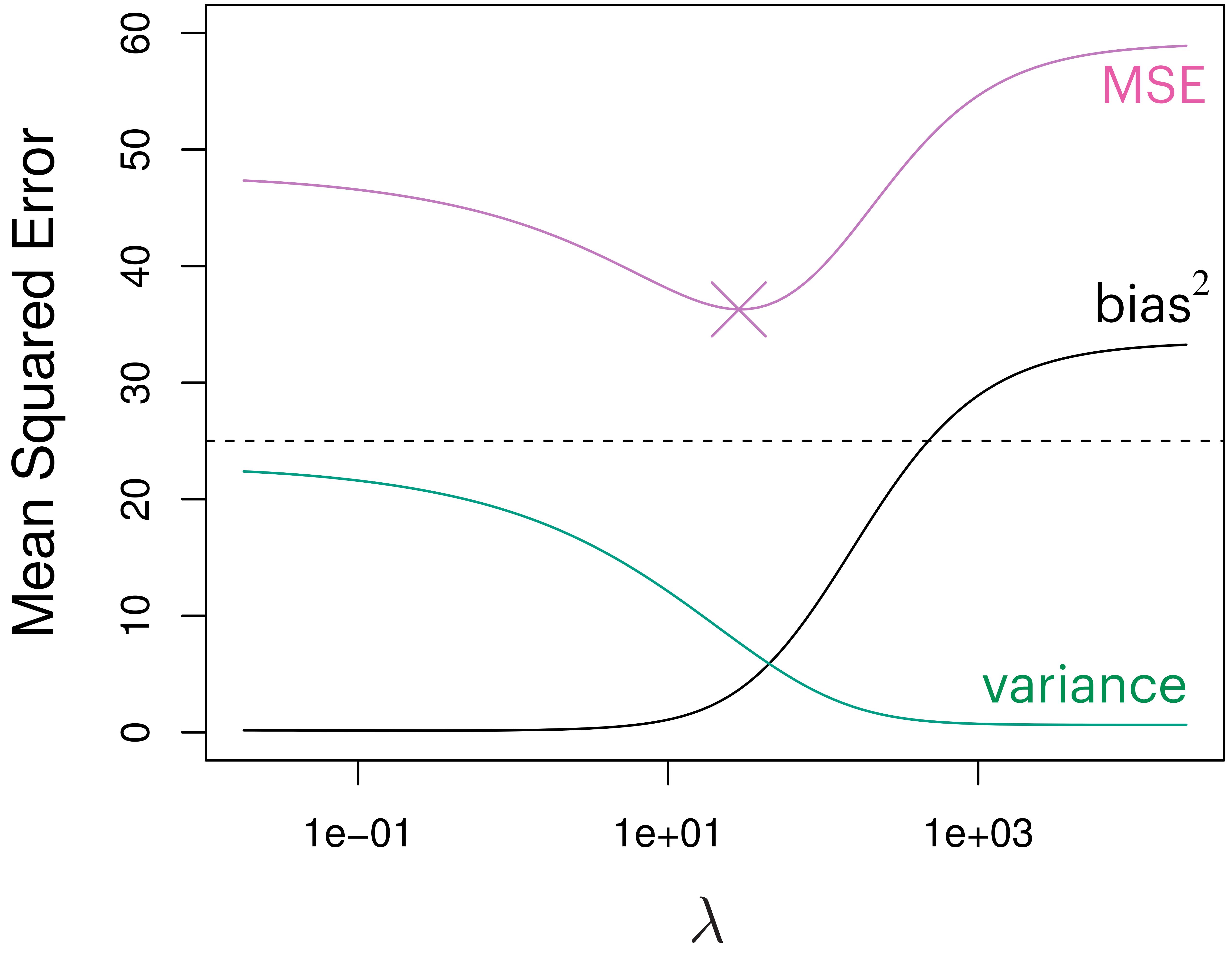
Ridge Regression

Bias-Variance Trade Off



[Source: ISLR2]

Lasso Regression

Least Absolute Shrinkage and Selection Operator

Least Squares produces estimates by minimizing

$$\text{RSS} = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \sum_{j-1}^{p} \hat{\beta}_{j_1} x_{ij})^2$$

Lasso regression instead minimizes

$$\underbrace{\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \sum_{j-1}^{p} \hat{\beta}_{j_1} x_{ij})^2}_{\text{model fit}} + \lambda \underbrace{\sum_{j=1}^{p} |\beta_j|}_{\text{penalty}} = \text{RSS} + \lambda \sum_{j=1}^{p} |\beta_j|$$

where $\lambda \geq 0$ is the tuning parameter controlling trade off between model fit and size of coefficients ($\lambda \to \infty, \hat{\beta}_j = 0$)

lasso uses ℓ_1 penalty