Linear Regression II

Lecture 3

Termeh Shafie

Assessing Model Fit

$$Y = f(X) + \epsilon$$

signal noise

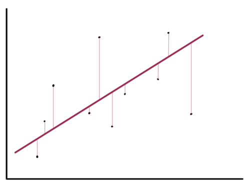
Loss Function

a metric for model performance, lower values are better

(for now we pretend that we have never heard of or seen cross-validation)

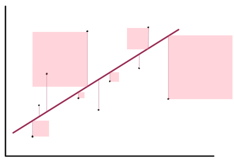
Assessing Model Fit

$$\text{MAE} = \frac{1}{n}\sum_i |\text{actual}_i - \text{predicted}_i|$$



Assessing Model Fit

$$\text{MSE} = \frac{1}{n}\sum_i (\text{actual}_i - \text{predicted}_i)^2$$

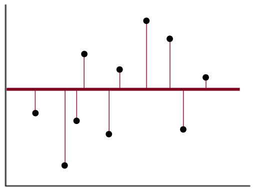


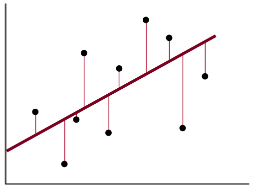
$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_i (\text{actual}_i - \text{predicted}_i)^2}$$

...what about a measure that is always on the same scale?

Assessing Model Fit

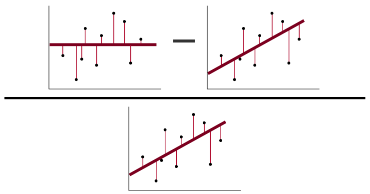
$$R^2 = 1 - \frac{\sum_i (\text{actual}_i - \text{predicted}_i)^2}{\sum_i (\text{actual}_i - \text{average})^2}$$





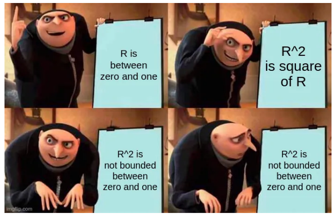
Assessing Model Fit

$$R^2 = 1 - \frac{\sum_i (\text{actual}_i - \text{predicted}_i)^2}{\sum_i (\text{actual}_i - \text{average})^2}$$



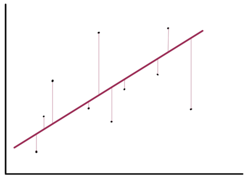
Assessing Model Fit

$$R^2 = 1 - \frac{\sum_i (\text{actual}_i - \text{predicted}_i)^2}{\sum_i (\text{actual}_i - \text{average})^2}$$



Assessing Model Fit

$$\text{MAPE} = \frac{1}{n}\sum_i \left| \frac{\text{actual}_i - \text{predicted}_i}{\text{actual}_i} \right|$$



Estimate $\hat{f}$ = Learn $\hat{f}$

$$Y = f(X) + \epsilon$$

sources of error:
irreducible error ϵ
reducible error $\hat{f}$

the squared error for a given estimate $\hat{f}$ is

$$E(\text{actual} - \text{predicted})^2 = E(Y - \hat{Y})^2$$

which factors as

$$E[f(X) + \epsilon - \hat{f}(X)]^2$$

$$\underbrace{[f(X) - \hat{f}(X)^2]}_{\text{reducible}} + \underbrace{\text{Var}(\epsilon)}_{\text{irreducible}}$$

until now, training data was the only data we considered
we compute reducible error (or MSE) on the same data used to learn $\hat{f}$
let's change that!

Training

training data set

$$\{(y_1, x_1,), \ldots, (y_n, x_n))\}$$

used to find function q that minimizes **Training MSE**

$$\hat{f} = \arg\min_q MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - q(x_i))^2$$

Testing

testing data sets (unseen)

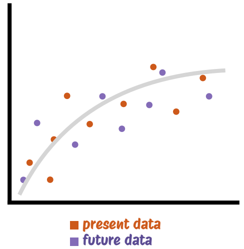
$$(y_0, x_0)$$

used to compute **Test MSE**

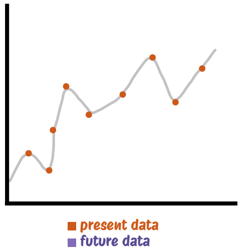
$$E[y_0 - \hat{f}(x_0)^2]$$

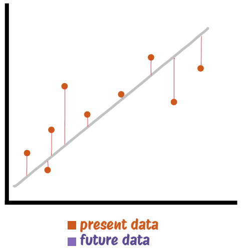
often not so closely related

Training and Testing

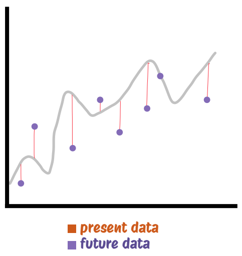


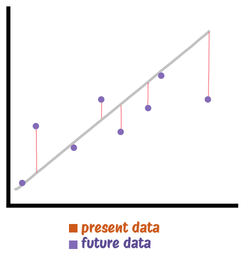
Training and Testing



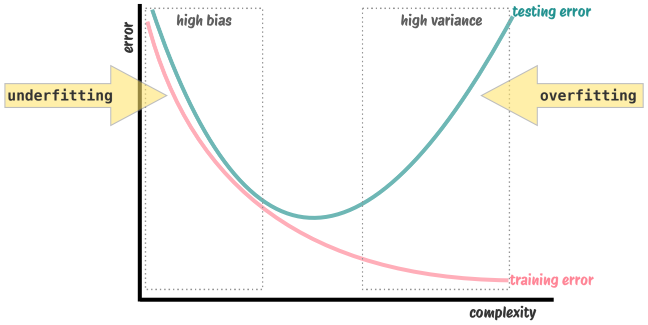


Training and Testing

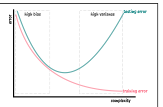




Bias Variance Trade-Off



Formalizing Bias Variance Trade-Off

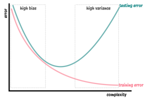


Expected test MSE

$$\underbrace{E\left(y_0 - \hat{f}(x_0)\right)^2}_{\text{expected MSE at } x_0 \text{ if we repeatedly estimated } f(x) \text{ with different training sets}} = \text{Var}(\hat{f}(x_0)) + \left[\text{bias}(\hat{f}(x_0))\right]^2 + \underbrace{\text{Var}(\epsilon)}_{\text{irreducible error}}$$

[try it out: https://floswald.shinyapps.io/bias_variance/]

Formalizing Bias Variance Trade-Off

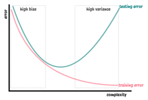


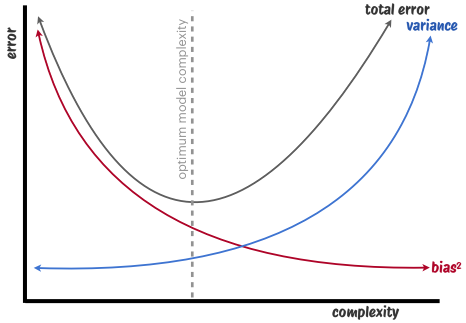
Expected **test MSE**

$$E\left(y_0 - \hat{f}(x_0)\right)^2 = \underbrace{\text{Var}(\hat{f}(x_0))}_{\text{variance increases with complexity}} + \underbrace{\left[\text{bias}(\hat{f}(x_0))\right]^2}_{\text{bias decreases with complexity}} + \text{Var}(\epsilon)$$

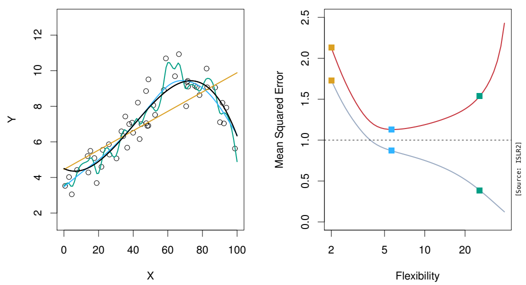
[try it out: https://floswald.shinyapps.io/bias_variance/]

Bias Variance Trade-Off

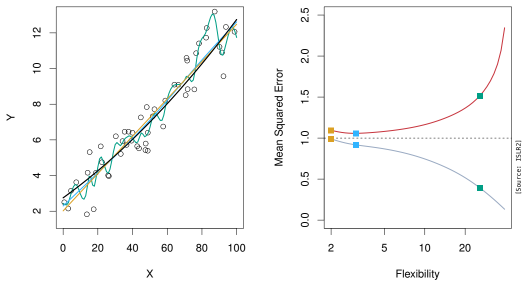




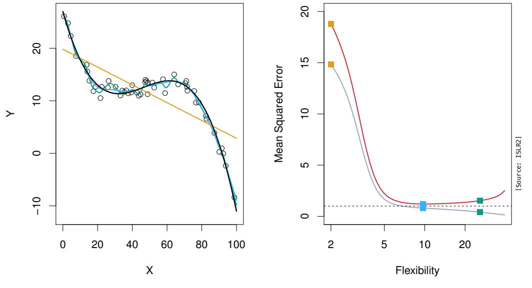
Example (a)



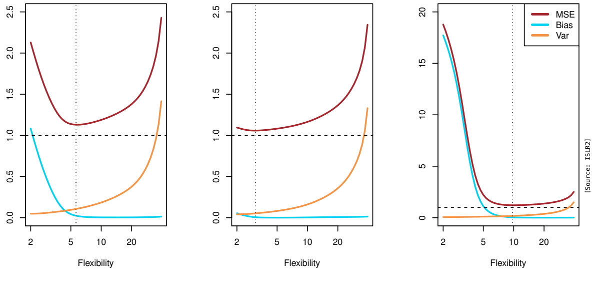
Example (b)



Example (c)



Example (a)-(c): Bias Variance Trade-Off



A Simulation Example

Estimate the conditional mean of Y given $\boldsymbol{X}$

$$Y = f(\boldsymbol{X}) + \epsilon$$

Assume probability model:

$$Y = 1 - 2x - 3x^2 + 5x^3 + \epsilon$$

where $\epsilon \sim \mathrm{N}(0,\sigma^2)$

Alternatively:

$$Y \mid X \sim \mathrm{N}(1 - 2x - 3x^2 + 5x^3, \sigma^2) \quad \text{or}$$

$$\mu(x) = E[Y \mid \boldsymbol{X} = \boldsymbol{x}] = 1 - 2x - 3x^2 + 5x^3$$

conditional mean is a **linear combination of the feature variables**

note: the true probability model and thus also $\mu(x)$ are often not known!

A Simulation Example

1. Simulate data from assumed probability model

$$Y = 1 - 2x - 3x^2 + 5x^3 + \epsilon$$

2. Fit three models to data:

   I. Degree 1 Polynomial $\mu(x) = \beta_0 + \beta_1 x$

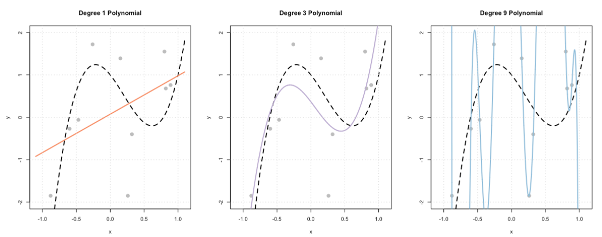
   II. Degree 3 Polynomial $\mu(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$

   III. Degree 9 Polynomial $\mu(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \ldots + \beta_9 x^9$

ENTER

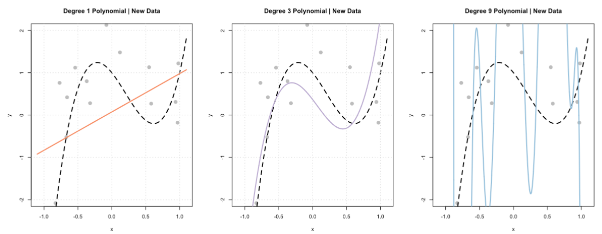
A Simulation Example

- How close is the estimated regression (mean) function to the data?
- How close is the estimated regression (mean) function to the true regression (mean) function?



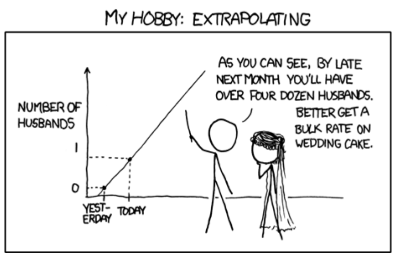
A Simulation Example

Generate new data and check...

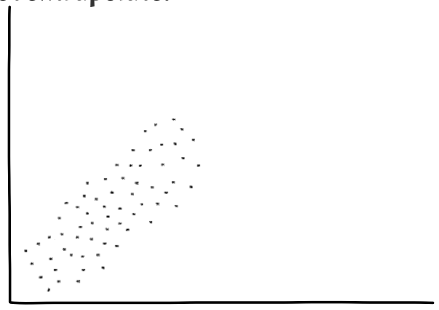


Some Do Not's!

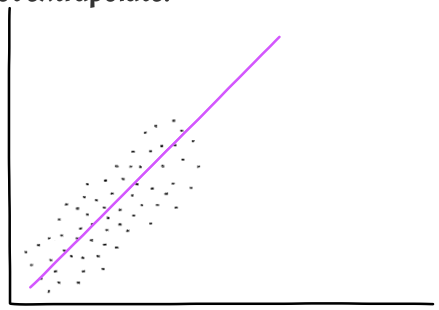
Extrapolation



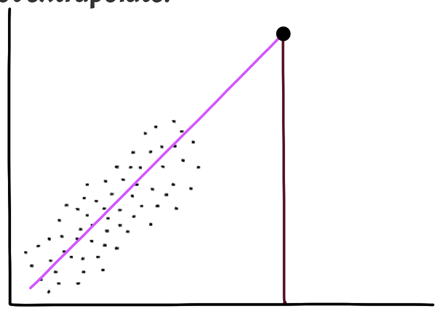
Do not extrapolate!



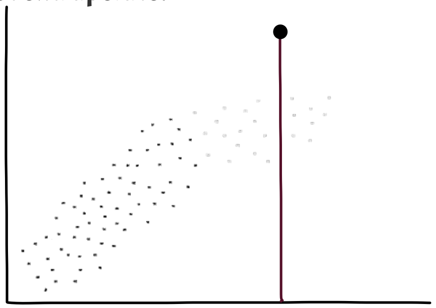
Do not extrapolate!



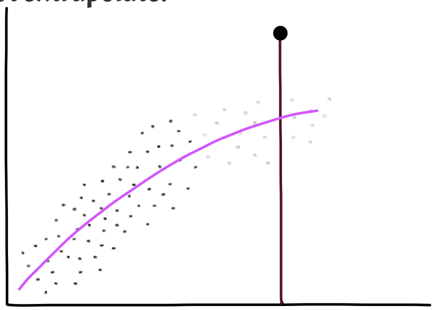
Do not extrapolate!



Do not extrapolate!



Do not extrapolate!



Do not fit model on test data!

In addition to the train-test split,
we will later split the data into validation set

This Week's Practical

Linear Regression: Evaluate Model Performance