Project 2

Chloe Chen & Kangrui Liu

9/12/2023

# 1. JWHT Chapter 2. Exercise 5.

## What are the advantages and disadvantages of a very flexible (versus a less flexible) approach for regression or classification?

### High Flexibility Advantages(less flexible - opposite of flexible):

1. High Predictive Capacity: Can capture complex patterns and non-linear relationships in data, which may more accurately model the true function.
2. Effective for Complex Data: Suitable for complex datasets with intricate relationships.
3. Overfitting Control: Overfitting can be mitigated with proper regularization techniques.

### High Flexibility Disadvantages(less flexible - opposite of flexible):

1. Complexity and Interpretability: Often challenging to interpret due to a large number of parameters.
2. Data Requirements: Requires substantial data for accurate parameter estimation.
3. Computational Complexity: May be computationally intensive.

## Under what circumstances might a more flexible approach be preferred to a less flexible approach?

1. The dataset is large and diverse with different groups, providing ample data for model training.
2. The relationship between predictors and the response variable is complex and non-linear.
3. Achieving the highest predictive accuracy is a top priority.
4. Adequate regularization techniques can be applied to control overfitting.
5. Computational resources are available for training and inference.

## When might a less flexible approach be preferred?

1. The dataset is small or has limited variability, making overfitting a concern.
2. Interpretability of the model is essential for understanding the relationships between variables.
3. Computational resources are limited or real-time processing is required.
4. The underlying data relationships are believed to be relatively simple and linear.
5. The model needs to be robust and less sensitive to outliers or noisy data.

# 2. Faraway Chapter 2. Exercise 2.

## The dataset uswages is drawn as a sample from the Current Population Survey in 1988.

## Fit a model with weekly wages as the response and years of education and experience as predictors in linear regression.

library(faraway)  
data("uswages")  
pred\_wage<-lm(wage~educ+exper,uswages)  
summary(pred\_wage)

##   
## Call:  
## lm(formula = wage ~ educ + exper, data = uswages)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1018.2 -237.9 -50.9 149.9 7228.6   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -242.7994 50.6816 -4.791 1.78e-06 \*\*\*  
## educ 51.1753 3.3419 15.313 < 2e-16 \*\*\*  
## exper 9.7748 0.7506 13.023 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 427.9 on 1997 degrees of freedom  
## Multiple R-squared: 0.1351, Adjusted R-squared: 0.1343   
## F-statistic: 156 on 2 and 1997 DF, p-value: < 2.2e-16

## Report and give a simple interpretation to the regression coefficient for years of education.