Math 133: Statistical Learning Methods - Module 2 Assessment

Pranav Jayakumar

February 19, 2025

1 Exercises

1.1 Election Predictions

We will use voter demographic data from the 2016 United States presidential elections to predict whether or not a person voted for Donald Trump.

1.1.1 Linear Model

We will begin by fitting a linear model of the form:

$$\hat{y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

Where:

1. $\hat{y}$ is `trump.vote`
2. β is the model coefficients
3. X_1 is `non.citizen`
4. X_2 is `unemployed`
5. X_3 is `metro`
6. ϵ is the residual error

```
data <- read.csv("../../data/election2016.csv")
trump_lm <- lm(trump.vote~non.citizen+unemployed+metro, data=data
    )
summary(trump_lm)
```

Table 1: Coefficients

Coefficients	Estimate	Std. Error	t-value	$\Pr(>\|t\|)$
(Intercept)	0.62064	0.06793	9.136	6.65×10^{-12}
non.citizen	-1.25402	0.57049	-2.198	0.0330
unemployed	1.58516	1.15351	1.374	0.1760
metro	-0.18211	0.10118	-1.800	0.0784

Table 2: Model Performance

Residual Standard Error	0.08011 on 46 degrees of freedom
Multiple R-squared	0.4072
Adjusted R-squared	0.3686
F-statistic	10.53 on 3 and 46 DF
p-value	2.157×10^{-5}

We observe that only the `non.citizen` variable is significant ($p = 0.0330$).

1.1.2 Reduced Linear Model

We will now fit a reduced linear model of the form:

$$\hat{y} = \beta_0 + \beta_1 X_1 + \epsilon$$

Where:

1. $\hat{y}$ remains as `trump.vote`
2. β is the model coefficients
3. X_1 is `non.citizen`
4. ϵ remains the residual error

```
reduced_lm <- update(trump_lm, .~.-unemployed-metro)
summary(reduced_lm)
```

Table 3: Reduced Model Coefficients

Coefficients	Estimate	Std. Error	t-value	$\Pr(>\|t\|)$
(Intercept)	0.5979	0.0225	26.569	$< 2 \times 10^{-16}$
non.citizen	-1.9392	0.3783	-5.126	5.25×10^{-6}