PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY *of* WASHINGTON

Introduction to Machine Learning

MLEARN 510A – Lesson 2

Logistics

- Some clarifications about the files attached to course modules
- Lesson X Knowledge Check: NOT graded
- Lesson X Code Talk: Discussion forum to discuss thoughts on that week’s lecture. Contributes 10% towards final grade
- Lesson X Assignment: Due the following week before next lecture. Contributes 90% towards final grade

Recap of Lesson 1

- What is Machine Learning?
  - Task **T**, Experience **E**, Performance **P**
- Three types of Machine Learning
  - Supervised, Unsupervised, Reinforcement
- Supervised learning
  - Classification, Regression
- Performance measures
  - Accuracy, Precision, Recall, ROC/AUC, MSE, MAE
- Model generalization
  - Underfitting, Overfitting, Regularization
- Parametric/non-parametric methods for statistical learning

Outline for Lecture 2

- ➢ Principle of Maximum Likelihood Estimation
- ➢ Simple Linear Regression
- ➢ Multiple Linear Regression
- ➢ Important Questions Related to a Regression Fit
- ➢ Handling Qualitative Predictors
- ➢ Linear Regression Diagnostics

Maximum Likelihood Estimation

- The Maximum Likelihood Estimation (MLE) is a method of estimating the parameters of a model

- Maximum Likelihood Estimation (MLE) is probably the most widely used method of estimation in statistics

- MLE has a strong intuitive appeal and often yields a reasonable estimator of θ

- Furthermore, if the sample is large, the method will yield an excellent estimator of θ

Maximum Likelihood Estimation

- Suppose that the random variables $X_1, \cdot\cdot\cdot, X_n$ form a random sample from a distribution f(x|θ); if X is continuous random variable, f(x|θ) is pdf, if X is discrete random variable, f(x|θ) is point mass function

- If observations are independent, $f(x_1, \cdot\cdot\cdot, x_n|\theta) = f(x_1|\theta)\cdot \cdot\cdot f(x_n|\theta)$

- Call $f(x_1, \cdot\cdot\cdot, x_n|\theta)$ as the *likelihood function*

- When the sample comes from a continuous distribution, it would again be natural to try to find a value of θ for which the probability density $f(x_1, \cdot\cdot\cdot, x_n|\theta)$ is large, and to use this value as an estimate of θ

Mathematical Form of MLE

- *We choose the parameter that maximizes the likelihood of having the obtained data at hand. With discrete distributions, the likelihood is the same as the probability. We choose the parameter for the density that maximizes the probability of the data coming from it*

- MLE requires us to maximize the likelihood function L(θ) with respect to the unknown parameter θ

$$l(\theta) = \log L(\theta) = \log \prod_{i=1}^{n} f(X_i|\theta) = \sum_{i=1}^{n} \log f(X_i|\theta)$$

- Maximizing *l*(θ) with respect to θ will give us the MLE estimator

MLE Example – Gaussian PDF

- Suppose we have two unknown parameters, μ and σ, therefore the parameter θ = (μ, σ) is a vector

$$l(\mu,\sigma) = \sum_{i=1}^{n}\left[-\log\sigma - \frac{1}{2}\log 2\pi - \frac{1}{2\sigma^2}(X_i-\mu)^2\right]$$

$$= -n\log\sigma - \frac{n}{2}\log 2\pi - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$$

- Setting the partial derivative to be 0, we have

$$\frac{\partial l(\mu,\sigma)}{\partial\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu) = 0$$

$$\frac{\partial l(\mu,\sigma)}{\partial\sigma} = -\frac{n}{\sigma} + \sigma^{-3}\sum_{i=1}^{n}(X_i-\mu)^2 = 0$$

Exercise

> Verify that the MLE is given by

$$\hat{\mu} = \overline{X} \quad \text{and} \quad \hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

Simple Linear Regression

- A straightforward linear approach for predicting a quantitative response

- Mathematically: $Y \approx \beta_0 + \beta_1 X$

  Where: "≈" means "approximately modeled as"

- Suppose we would like to predict sales based on TV advertising

$$\texttt{sales} \approx \beta_0 + \beta_1 \times \texttt{TV}$$

  Where: β_0 – intercept, β_1 – slope (coefficients)

Estimating Coefficients

- We can use training data to estimate the coefficients:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

where: ‘^’ means ‘estimated value’

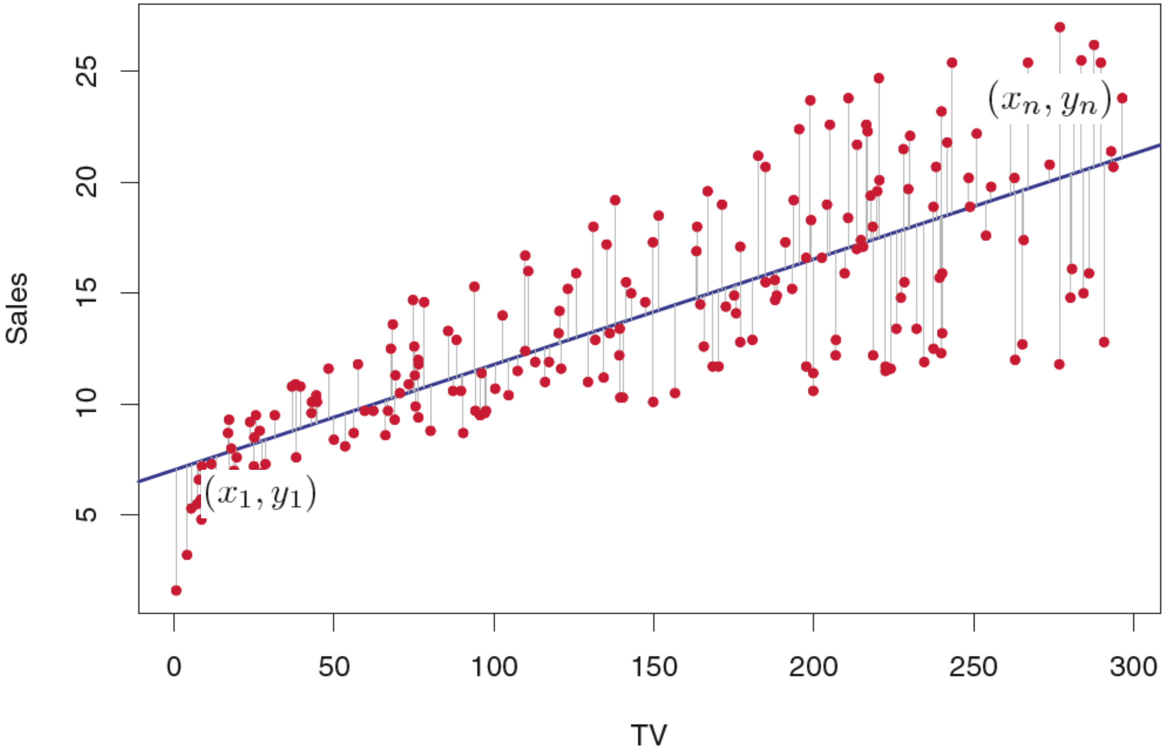
- We will do it by fitting this equation to the model:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

where: ‘ε’ is the error term

Estimating Coefficients

$$(x_1, y_1),\ (x_2, y_2), \ldots,\ (x_n, y_n)$$



Minimizing Residuals

- Residual is the difference between actual and predicted value

- Residual for the ith observation is given by

$$e_i = y_i - \hat{y}_i$$

- Residual sum of squares

$$\mathrm{RSS} = e_1^2 + e_2^2 + \cdots + e_n^2 \; = \; \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$\mathrm{RSS} = (y_1 - \hat{\beta}_0 - \hat{\beta}_1 x_1)^2 + (y_2 - \hat{\beta}_0 - \hat{\beta}_1 x_2)^2 + \ldots + (y_n - \hat{\beta}_0 - \hat{\beta}_1 x_n)^2$$

- Residual Standard Error

$$\sqrt{\mathrm{RSS}/(n-2)}$$

Quiz

- Recall your high school calculus class. How do we find the optimal values of coefficients which minimize RSS?

$$\mathrm{RSS} = e_1^2 + e_2^2 + \cdots + e_n^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$\mathrm{RSS} = (y_1 - \hat{\beta}_0 - \hat{\beta}_1 x_1)^2 + (y_2 - \hat{\beta}_0 - \hat{\beta}_1 x_2)^2 + \ldots + (y_n - \hat{\beta}_0 - \hat{\beta}_1 x_n)^2$$

Estimated Coefficients

- Simple linear regression only has one predictor
- Slope and intercept are computed as:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Standard Error of Regression Coefficients

- The standard error of a regression coefficient quantifies our uncertainty about the regression coefficient

- A 95% confidence interval is defined as a range of values such that with 95% probability, the range will contain the true unknown value of the parameter

$$\mathrm{SE}(\hat{\beta}_0)^2 = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \right] \qquad \mathrm{SE}(\hat{\beta}_1)^2 = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Hypothesis Test for a Regression Coefficient

Null hypothesis

$$H_0 : \beta_1 = 0$$

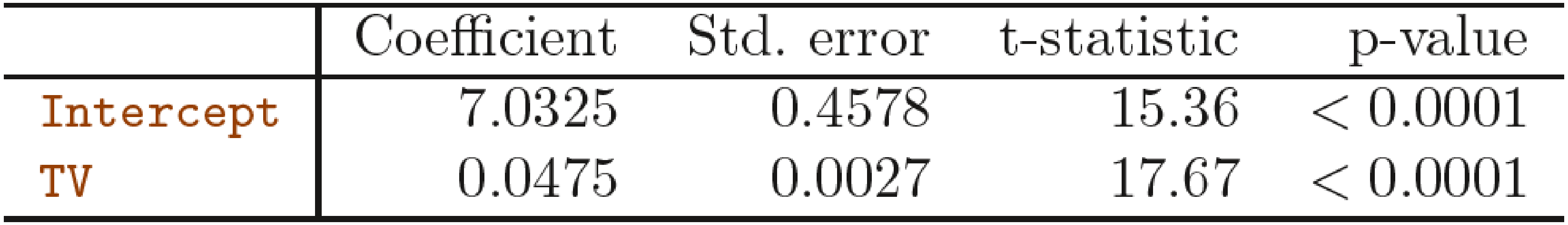
Alternative hypothesis

$$H_a : \beta_1 \neq 0$$

- Test Statistic: the ratio of a difference to its standard error

$$t = \frac{\hat{\beta}_1 - 0}{\mathrm{SE}(\hat{\beta}_1)}$$

Evaluating Coefficients



	Coefficient	Std. error	t-statistic	p-value
Intercept	7.0325	0.4578	15.36	< 0.0001
TV	0.0475	0.0027	17.67	< 0.0001

Residual Standard Error

- Considered a measure of the “lack of fit” of the model

- It can be thought of as ‘average deviation’ of the model from the actual data point

$$\mathrm{RSE} = \sqrt{\frac{1}{n-2}\mathrm{RSS}} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}.$$

Coefficient of Determination – R²

- Provides an absolute measure of lack of fit of the model (unlike RSE)
- Measures the proportion of variability in Y that can be explained using X
- Accepts value from 0 (no fit) to 1 (perfect fit)

$$\text{TSS} = \sum (y_i - \bar{y})^2$$

$$\text{RSS} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$R^2 = \frac{\text{TSS} - \text{RSS}}{\text{TSS}} = 1 - \frac{\text{RSS}}{\text{TSS}}$$

Pearson Correlation Coefficient vs R²

$$\mathrm{Cor}(X,Y) = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}}$$

- In simple linear models, $r^2 = Cor(X,Y)^2 = R^2$
- Problematic interpretation of correlation coefficient when multiple variables present

Multiple Linear Regression

- Multiple - more than one predictor fitted at the same time:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \epsilon$$

- For the advertising dataset:

$$\texttt{sales} = \beta_0 + \beta_1 \times \texttt{TV} + \beta_2 \times \texttt{radio} + \beta_3 \times \texttt{newspaper} + \epsilon$$

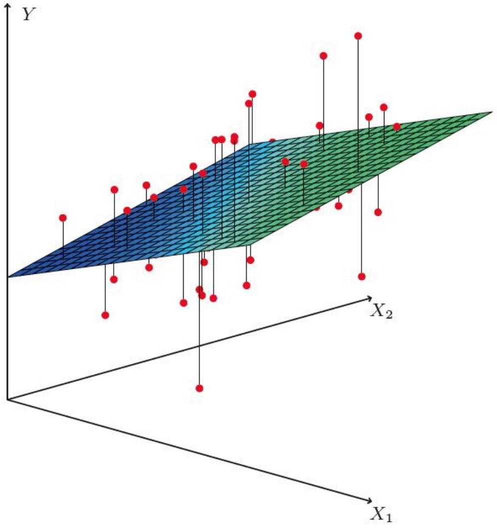
- Thus, our general equation for estimating the coefficients is:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_p x_p$$

Multiple Linear Regression

- Estimate the coefficients in equation

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_p x_p$$



- By minimizing

$$\begin{aligned} \mathrm{RSS} &= \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \\ &= \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_p x_{ip})^2 \end{aligned}$$

Multiple Linear Regression Using MLE

- The observations can be expressed in matrix form as

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & & \\ \dots & & & \\ x_{n1} & & & x_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \dots \\ \epsilon_n \end{bmatrix}$$

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

- The noise distribution is assumed as $\mathcal{N}(\mathbf{0}, \boldsymbol{I})$. The conditional distribution is expressed as

$$p(\boldsymbol{y}|\boldsymbol{X};\boldsymbol{\beta}) = \frac{1}{2\pi^{n/2}} \exp\left(\frac{-(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{2}\right)$$

Matrix Algebra – Multiple Linear Regression

- The log-likelihood can be expressed as

$$\log p(\boldsymbol{y}|\boldsymbol{X};\beta) \propto \frac{-(\boldsymbol{y}-\boldsymbol{X}\beta)^T(\boldsymbol{y}-\boldsymbol{X}\beta)}{2}$$

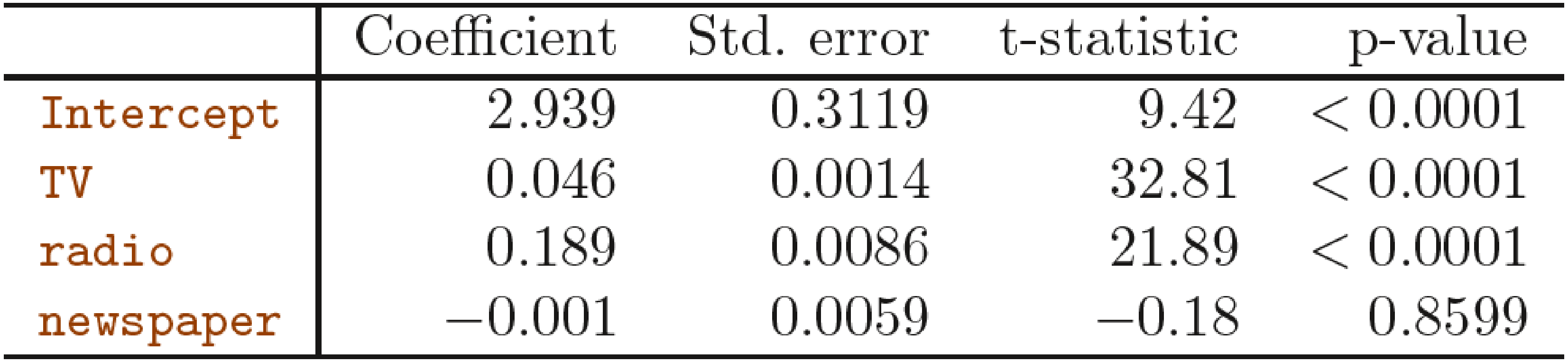
- Solving for coefficient vector gives

$$\hat{\beta} = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

- The covariance matrix is given by

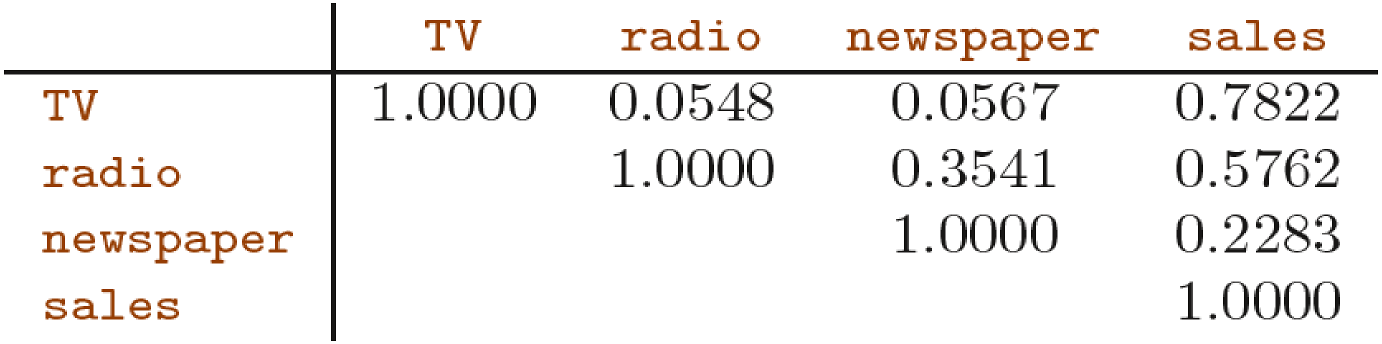
$$\mathrm{Cov}\left(\hat{\beta}\right) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}$$

Multiple Regression – Advertising Data



	Coefficient	Std. error	t-statistic	p-value
Intercept	2.939	0.3119	9.42	< 0.0001
TV	0.046	0.0014	32.81	< 0.0001
radio	0.189	0.0086	21.89	< 0.0001
newspaper	−0.001	0.0059	−0.18	0.8599

Correlation Matrix



	TV	radio	newspaper	sales
TV	1.0000	0.0548	0.0567	0.7822
radio		1.0000	0.3541	0.5762
newspaper			1.0000	0.2283
sales				1.0000

Quiz

- Suppose that you are working on a data science project. In your over-eagerness to show results, you make an error in your code. You accidentally copy the same data twice.
- Let’s say you started with 100 data samples. Your code copies the 100 data samples twice and you end up with 200 samples of data, where the last 100 are simply a copy of the first 100.

- Does anything change in Linear Regression? If yes, what changes? If not, why not?

$$\hat{\beta} = \left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{y}$$

$$\mathrm{Cov}\left(\hat{\beta}\right) = \left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1}$$

Some Important Questions

- Is at least one of the predictors X_1, X_2,…, X_p useful in predicting the response?
- 2. Do all the predictors help to explain Y, or is only a subset of the predictors useful?
- 3. How well does the model fit the data?
- 4. Given a set of predictor values, what response value should we predict and how accurate is our prediction?

Is There a Relationship Between Predictors and Response Variable?

- Perform hypothesis testing to answer this question:

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$$

$$H_a : \text{ at least one } \beta_j \text{ is non-zero}$$

- Specifically, calculate the F-statistics:

$$F = \frac{(\text{TSS} - \text{RSS})/p}{\text{RSS}/(n - p - 1)}$$

$$\text{TSS} = \sum (y_i - \bar{y})^2$$

$$\text{RSS} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

- If the null hypothesis is True, the ratio will be 1
- For our dataset, the value is 570

Is There a Relationship Between Predictors and Response Variable?

- ➢ Quality of the model can be evaluated using:
  - ❑ Akaike information criterion (AIC), Bayesian information criterion (BIC), and adjusted R^2
- ➢ For a model with p parameters, there are 2^p possible model combinations
- ➢ The smaller subset of optimal parameters (features) can be selected using:
  - ❑ **Forward Selection:** add one variable at a time, choosing the variable that best reduces the RSS
  - ❑ **Backward Selection:** remove one variable at a time, choosing the variable with the largest p value
  - ❑ **Mixed Selection:** use forward selection, but remove any variable that exceeds a threshold p value

Three Types of Uncertainty

- Three types of uncertainty:

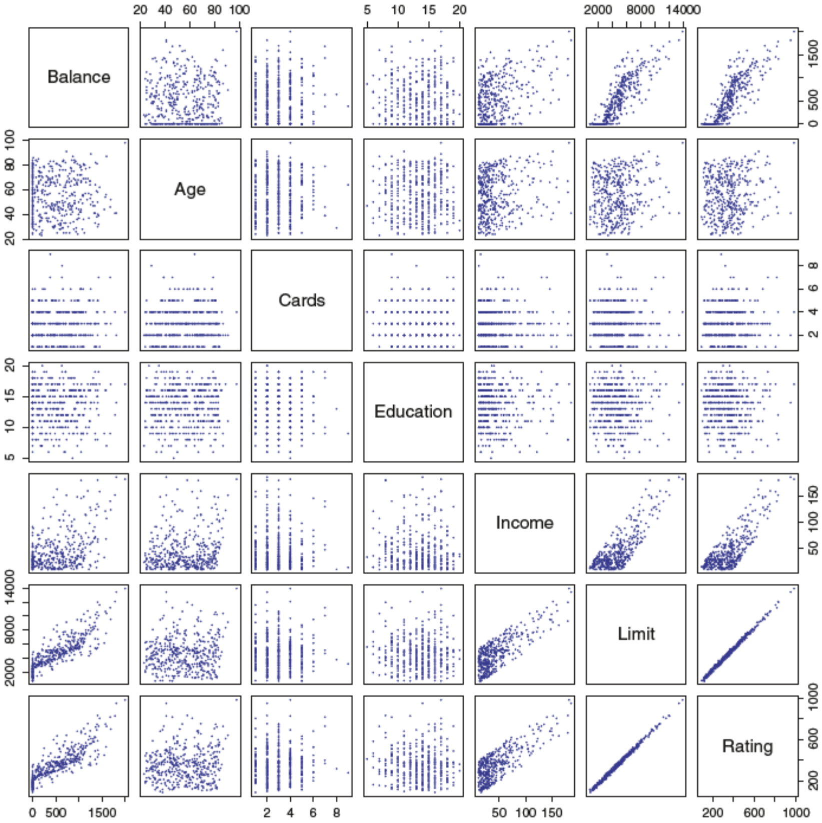
- **Confidence interval:** for the prediction of the mean output variable (the mean for a particular input vector)

- **Model bias:** the error caused by choosing a linear model when the true model [which is unknown] does not match the model used

- **Prediction interval:** for the prediction of the output variable

Other Considerations in Regression

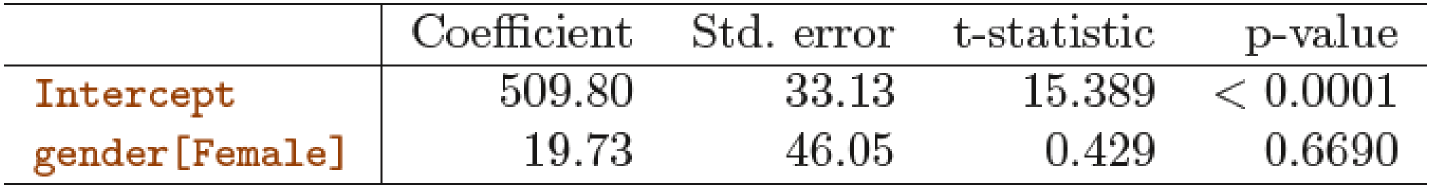
- Quantitative vs. qualitative predictors
- Two levels vs. multiple levels
- Non-linear relationships



Predictor With Only Two Levels

- Create indicator dummy variable x_i that denotes presence of a feature
- Interpretation: the average Balance for gender=Male is $509.80, while the average Balance for gender=Female is $19.73 more

$$x_i = \begin{cases} 1 & \text{if } i\text{th person is female} \\ 0 & \text{if } i\text{th person is male,} \end{cases} \qquad y_i = \beta_0 + \beta_1 x_i + \epsilon_i = \begin{cases} \beta_0 + \beta_1 + \epsilon_i & \text{if } i\text{th person is female} \\ \beta_0 + \epsilon_i & \text{if } i\text{th person is male.} \end{cases}$$



	Coefficient	Std. error	t-statistic	p-value
`Intercept`	509.80	33.13	15.389	< 0.0001
`gender[Female]`	19.73	46.05	0.429	0.6690

Alternative Coding Scheme

$$x_i = \begin{cases} 1 & \text{if } i\text{th person is female} \\ -1 & \text{if } i\text{th person is male} \end{cases}$$

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i = \begin{cases} \beta_0 + \beta_1 + \epsilon_i & \text{if } i\text{th person is female} \\ \beta_0 - \beta_1 + \epsilon_i & \text{if } i\text{th person is male.} \end{cases}$$

Interpretation:

Average overall balance is Beta[0], with:

- Beta[1] added to derive the average Balance for gender=Female
- Beta[1] subtracted to derive the average Balance for gender=Male

Qualitative Predictors with More than Two Values

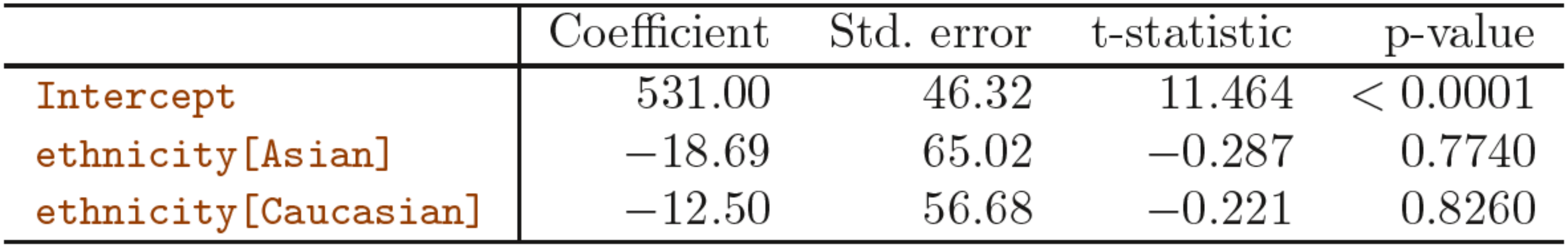
- What if a predictor has more than two levels?
- Let’s consider a predictor with 3 possible levels:
  - African American
  - Caucasian
  - Asian

Then:

$$x_{i1} = \begin{cases} 1 & \text{if } i\text{th person is Asian} \\ 0 & \text{if } i\text{th person is not Asian} \end{cases} \qquad x_{i2} = \begin{cases} 1 & \text{if } i\text{th person is Caucasian} \\ 0 & \text{if } i\text{th person is not Caucasian} \end{cases}$$

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i = \begin{cases} \beta_0 + \beta_1 + \epsilon_i & \text{if } i\text{th person is Asian} \\ \beta_0 + \beta_2 + \epsilon_i & \text{if } i\text{th person is Caucasian} \\ \beta_0 + \epsilon_i & \text{if } i\text{th person is African American.} \end{cases}$$

Evaluating Predictors



	Coefficient	Std. error	t-statistic	p-value
Intercept	531.00	46.32	11.464	< 0.0001
ethnicity[Asian]	-18.69	65.02	-0.287	0.7740
ethnicity[Caucasian]	-12.50	56.68	-0.221	0.8260

This F-test has a p-value of 0.96, indicating that we cannot reject the null hypothesis that there is no relationship between **balance** and **ethnicity**.

Predictor Interaction Effects

- When the effect on Y of increasing X_1 depends on another X_2.

- Example:
  - Maybe the effect on Salary (Y) when increasing Position (X_1) depends on gender (X_2)?
  - For example maybe Male salaries go up faster (or slower) than Females as they get promoted.

- Advertising example:
  - TV and radio advertising both increase sales.
  - Perhaps spending money on both may increase sales more than spending the same amount on one alone?

Interaction in Advertising Example

$$\boxed{Sales = b_0 + b_1 \times TV + b_2 \times Radio + b_3 \times TV \times Radio}$$

Interaction Term

$$\boxed{Sales = b_0 + (b_1 + b_3 \times Radio) \times TV + b_2 \times Radio}$$

- Spending $1 extra on TV increases average sales by 0.0191 + 0.0011Radio

$$\boxed{Sales = b_0 + (b_2 + b_3 \times TV) \times Radio + b_2 \times TV}$$

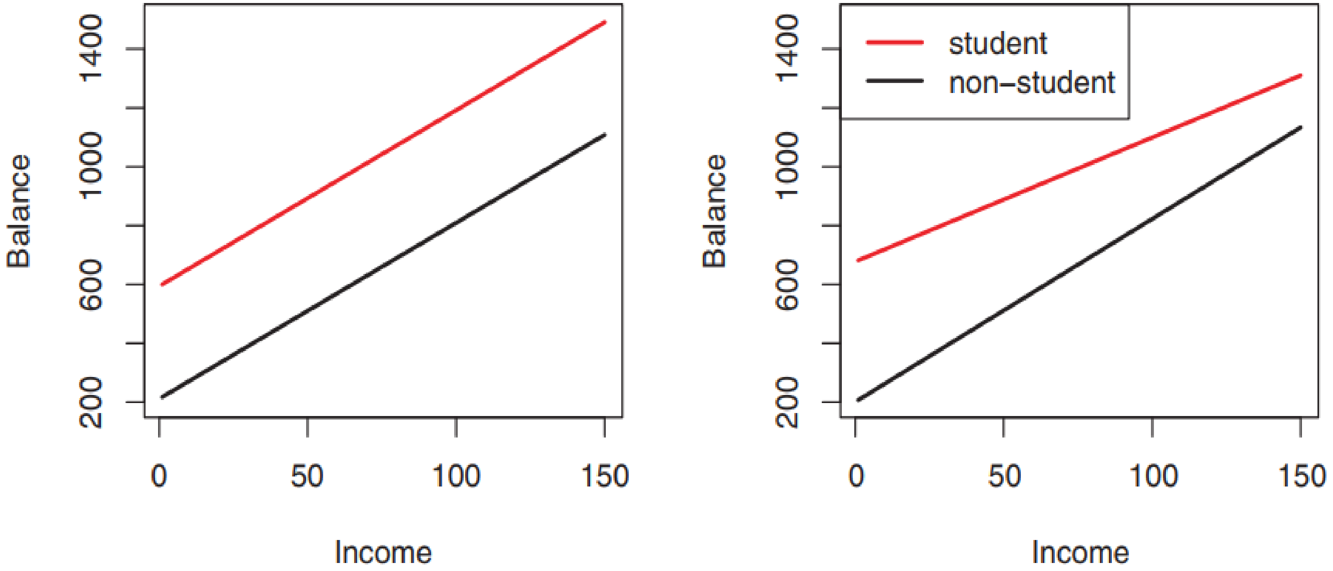
- Spending $1 extra on Radio increases average sales by 0.0289 + 0.0011TV

	Coefficient	Std. error	t-statistic	p-value
Intercept	6.7502	0.248	27.23	< 0.0001
TV	0.0191	0.002	12.70	< 0.0001
radio	0.0289	0.009	3.24	0.0014
TV×radio	0.0011	0.000	20.73	< 0.0001

Extensions – Qualitative Variable Interaction

$$
\begin{aligned}
\texttt{balance}_i &\approx \beta_0 + \beta_1 \times \texttt{income}_i + \begin{cases} \beta_2 & \text{if } i\text{th person is a student} \\ 0 & \text{if } i\text{th person is not a student} \end{cases} \\
&= \beta_1 \times \texttt{income}_i + \begin{cases} \beta_0 + \beta_2 & \text{if } i\text{th person is a student} \\ \beta_0 & \text{if } i\text{th person is not a student} \end{cases}
\end{aligned}
$$

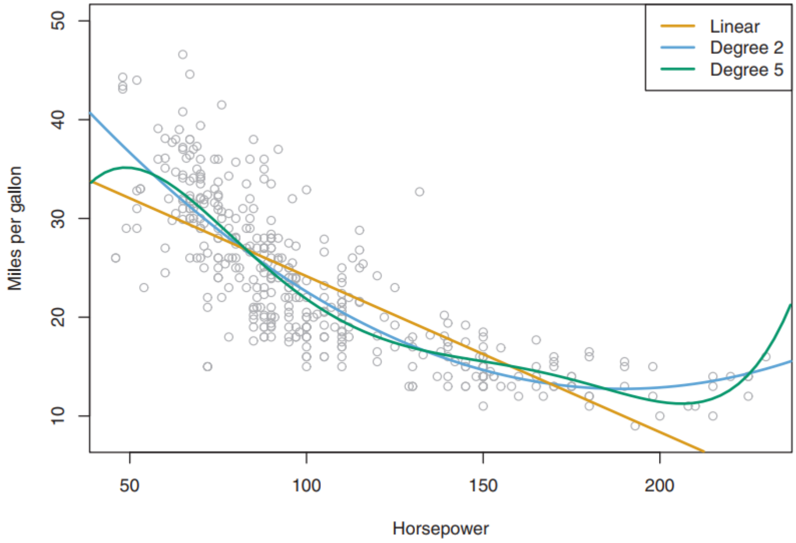
$$
\begin{aligned}
\texttt{balance}_i &\approx \beta_0 + \beta_1 \times \texttt{income}_i + \begin{cases} \beta_2 + \beta_3 \times \texttt{income}_i & \text{if student} \\ 0 & \text{if not student} \end{cases} \\
&= \begin{cases} (\beta_0 + \beta_2) + (\beta_1 + \beta_3) \times \texttt{income}_i & \text{if student} \\ \beta_0 + \beta_1 \times \texttt{income}_i & \text{if not student} \end{cases}
\end{aligned}
$$



Extensions – Non-Linear Relationships

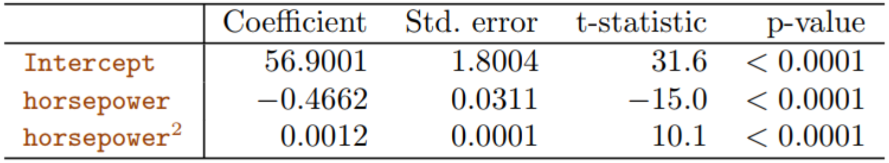
- Extend the model using polynomial regression:

$$\texttt{mpg} = \beta_0 + \beta_1 \times \texttt{horsepower} + \beta_2 \times \texttt{horsepower}^2 + \epsilon$$



- Model is still linear!

- It is possible to extend the model with higher-order terms (*hp*^{3}, *hp*^{4}, etc.)



	Coefficient	Std. error	t-statistic	p-value
`Intercept`	56.9001	1.8004	31.6	< 0.0001
`horsepower`	-0.4662	0.0311	-15.0	< 0.0001
`horsepower`^{2}	0.0012	0.0001	10.1	< 0.0001

- Always check that the additional terms are statistically significant

Potential Issues With Linear Regression

- There are several possible problems that one may encounter when fitting the linear regression model

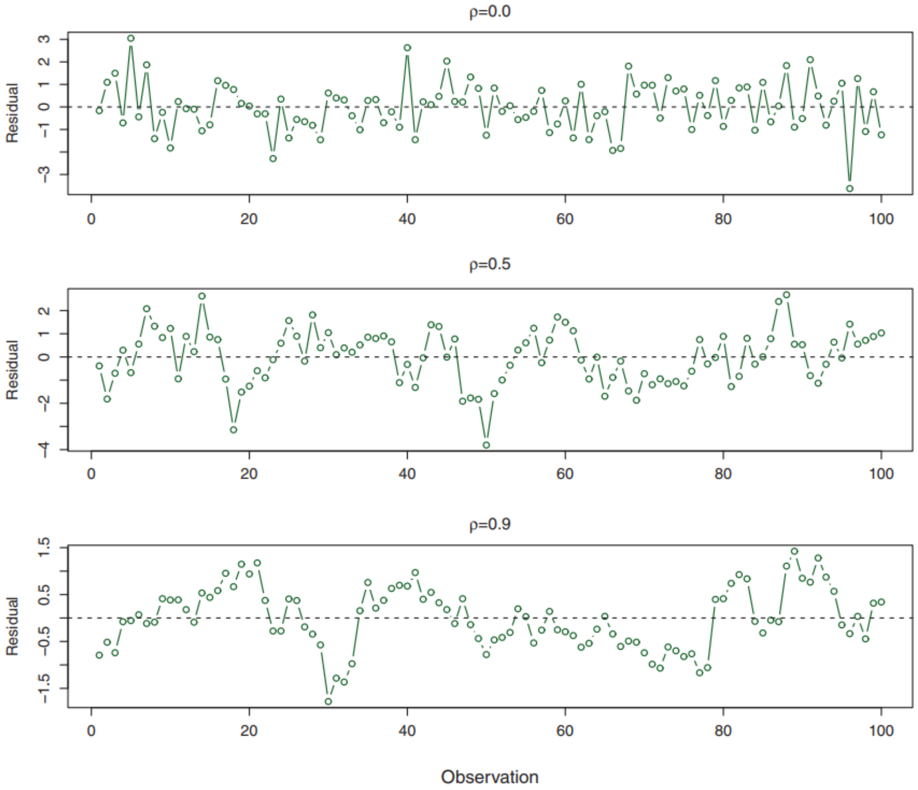
- Non-linearity of the data
- Dependence of the error terms
- Non-constant variance of error terms
- Outliers
- High leverage points
- Collinearity

Problem – Correlation in Error Terms

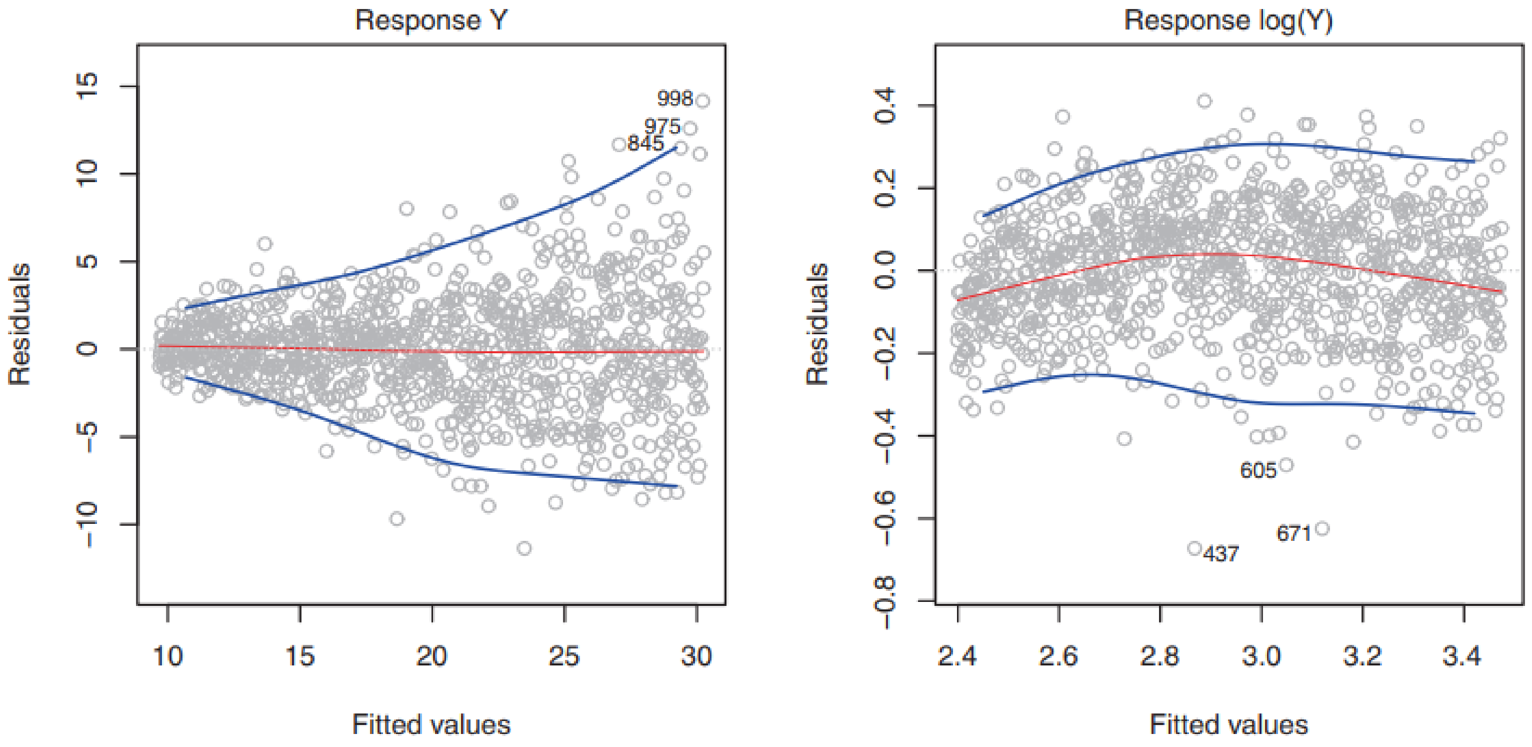
Problem: Can lead to underestimating the error terms (thus, confidence intervals will be narrower than they should be)

Diagnosis: Plot residuals of fitted values and observe "tracking"

Solution: Consider methods that work with time series



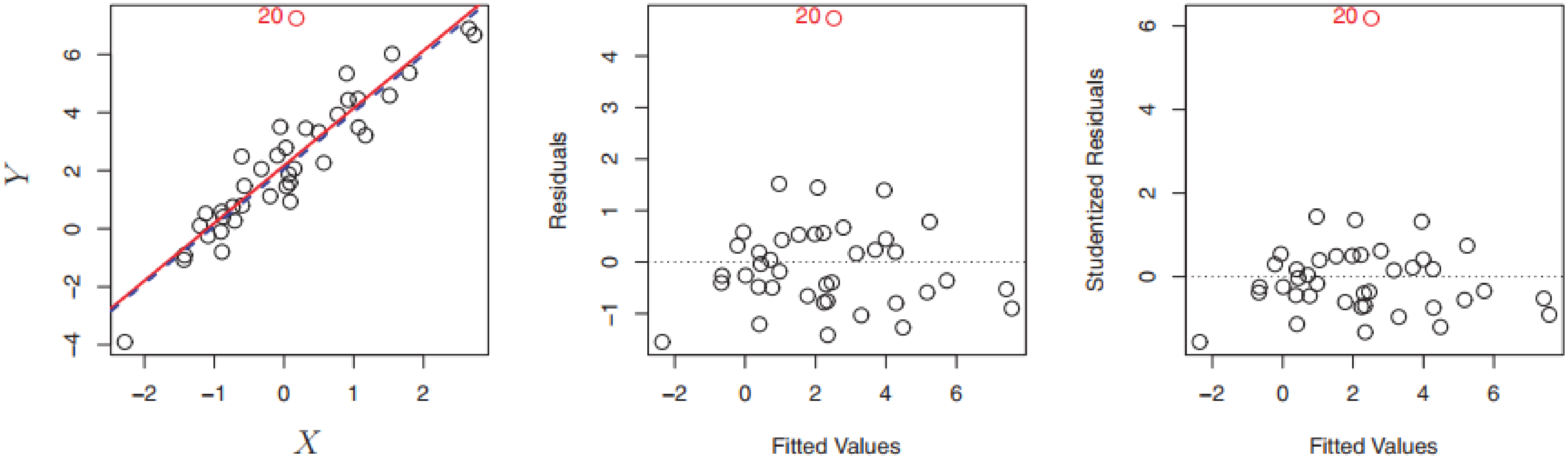
Problem: Non-Constant Variance of Error Terms (Heteroscedasticity)



Diagnosis: Plot residuals against fitted values

Solution : Consider transforming the output (e.g. *log(Y)*, √*(Y)*)

Problem: Outliers

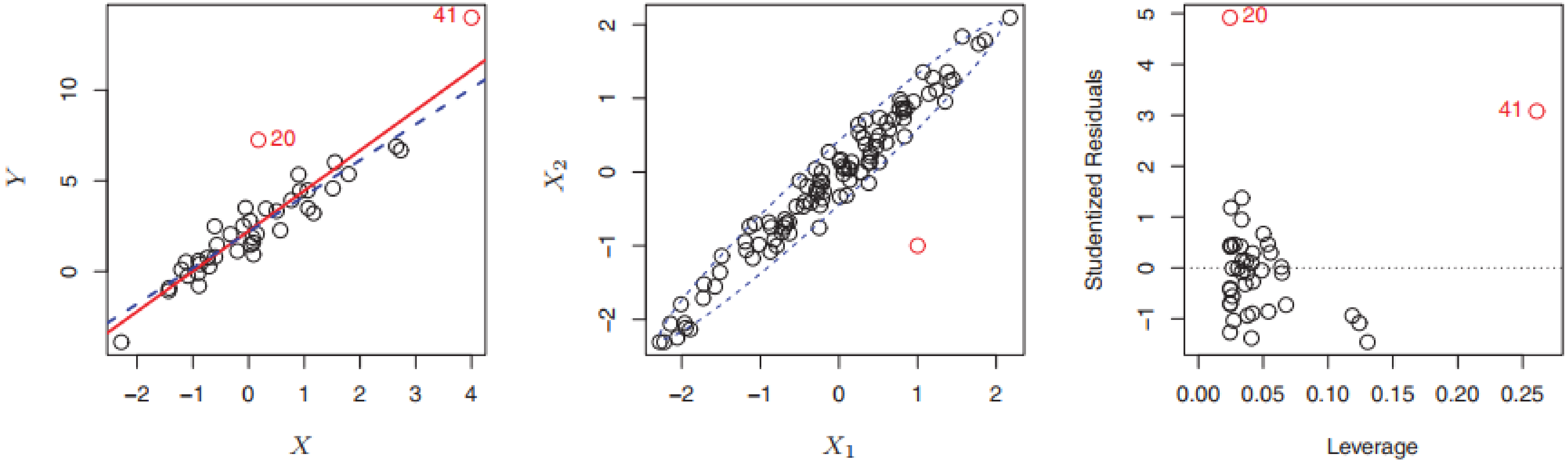


Problem: Unusual output value may increase the RSE and reduce R^2

Diagnosis: Plot residuals of fitted values (or Studentized residuals)

Solution: Remove the outliers, if they are due to data sampling error

Problem: High Leverage Points

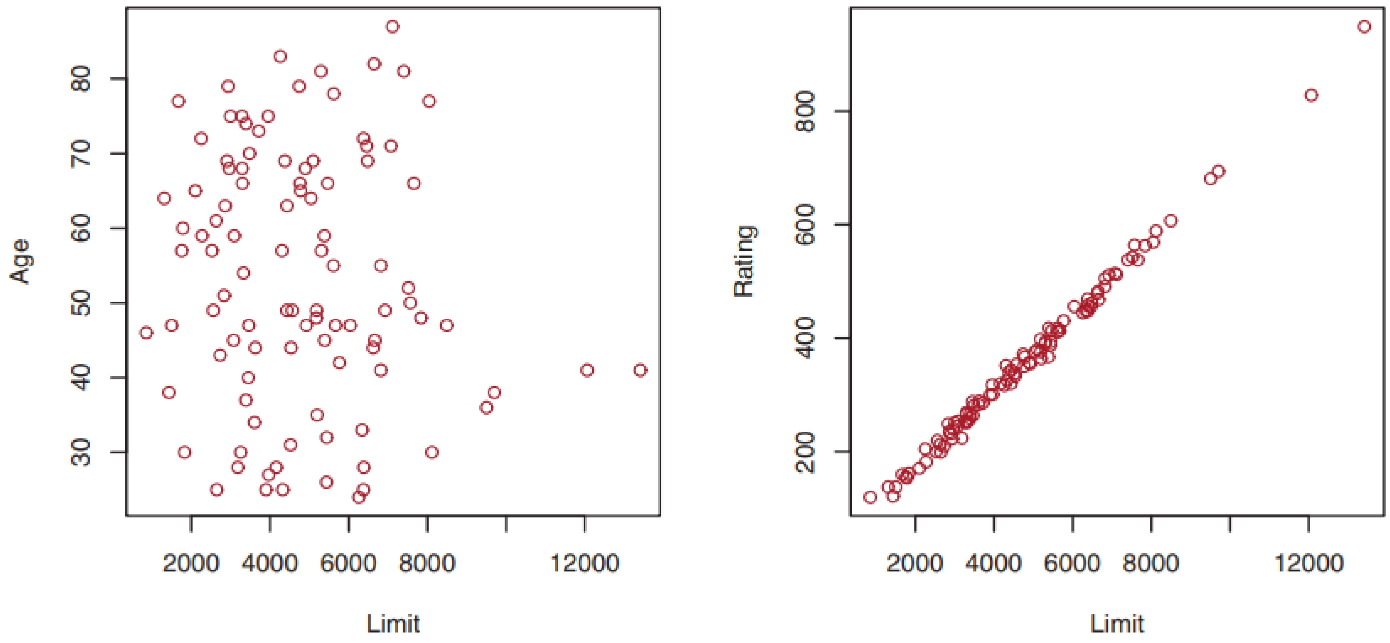


Problem: Unusual (high leverage) value for x_i

Diagnosis: Look for observations outside of the normal range or calculate leverage statistics:

Solution: Compare fits with/without high leverage points

Problem: Collinearity



Problem: Highly collinear variables make it difficult to separate the effects of collinear variables on the response

Diagnosis: Scatterplots / correlations of respective variables

Solution: ?

Problem: Collinearity

Simple:

- Correlation matrix

Better:

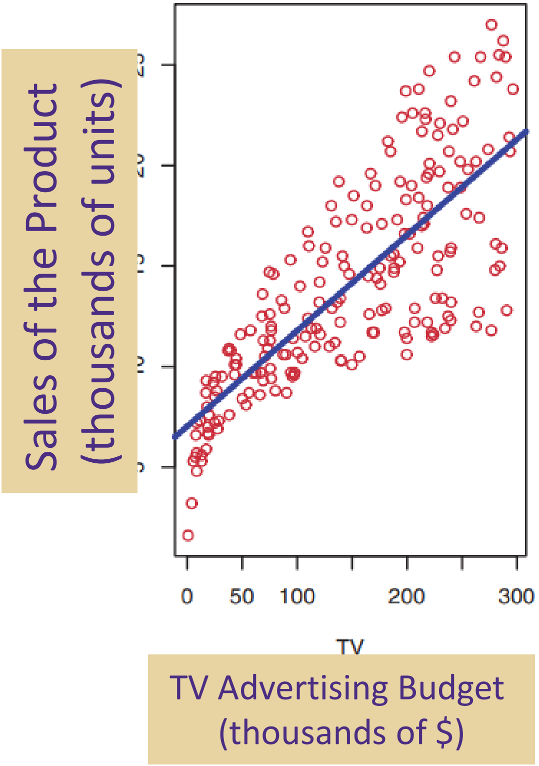
- Variance inflation factor (to detect multicollinearity):

$$\mathrm{VIF}(\hat{\beta}_j) = \frac{1}{1 - R^2_{X_j|X_{-j}}}$$

- Large values indicate collinearity problem

Return to Questions for Advertising Data

1. Is there a relationship between advertising budget and sales? → **F-test**
2. How strong is the relationship between advertising budget and sales? → **RSE, R^2**
3. Which media contributes to sales? → **t-test p-value**
4. How accurately can we estimate the effect of each medium on sales? → **Coefficient confidence intervals**
5. How accurately can we predict future sales? → **Prediction intervals**
6. Is the relationship linear? → **Residual plot**
7. Is there synergy among the advertising data? → **Interaction terms**



Linear Regression vs. K-nearest Neighbors

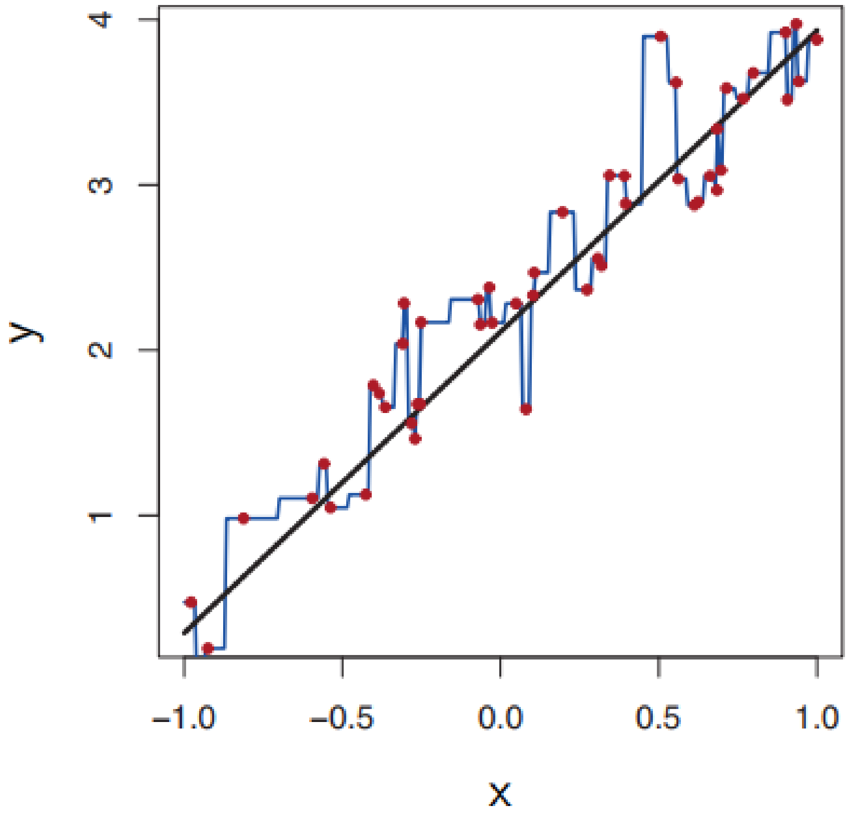
Linear Regression:

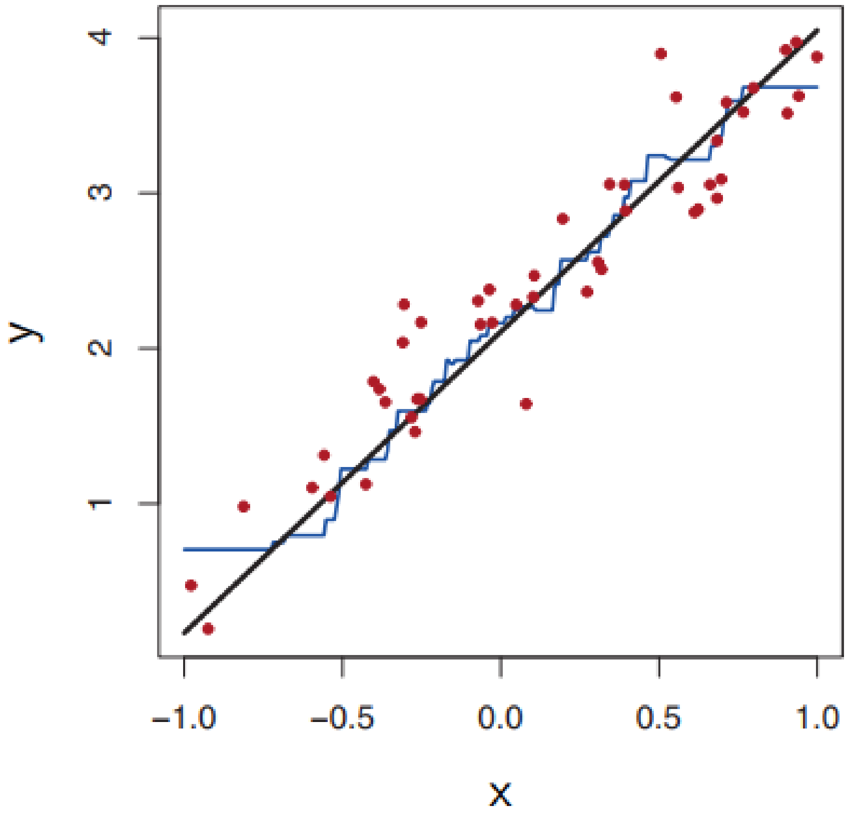
- Parametric
  - We assume an underlying functional form for *f(x)*
- Pros:
  - Easy to fit
  - Easy interpretation of coefficients
  - Easy to do significance testing
- Cons:
  - Poor performance, if wrong about the functional form

K-Nearest Neighbors:

- Non-parametric
  - No assumptions about the functional form for f(x)
- Pros:
  - No requirements for knowledge about underlying function
  - More flexible
- Cons:
  - Can mask underlying function
  - No extrapolation beyond available data

KNN Regression With Only 1 Predictor





ON-BRAND STATEMENT

FOR GENERAL USE

> What defines the students and faculty of the University of Washington? Above all, it’s our belief in possibility and our unshakable optimism. It’s a connection to others, both near and far. It’s a hunger that pushes us to tackle challenges and pursue progress. It’s the conviction that together we can create a world of good. And it’s our determination to Be Boundless. Join the journey at **uw.edu**.