Linear Regression

Project ENABLE

June 6, 2019

THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

Suggested Reading

- Witten, I. H., Frank, E., & Hall, M. A. (2005). Data Mining: Practical Machine Learning Tools and Techniques: Chapter 4.6. (uploaded to the Slack).

Machine Learning Algorithms

- **Machine Learning algorithm**: a procedure in developing computer programs that improve their performance with "experience."
- Types of machine learning: Naïve Bayes, Linear Regression, Support Vector Machine, Decision Tree, Neural Network, Deep Learning, and so on.
- Learning all algorithms are over the scope of this summer boot camp. We will focus on Linear Regression and Naïve Bayes as examples of Machine Learning.

Linear Regression

- When the outcome as well as all attributes are numeric, linear regression is a natural technique.
- Equation:

$$y = w_0 + \sum_{j=1}^{n} w_j x_j$$

where:

  - x_j is the attribute values
  - w_j are weights

Linear Regression (cont …)

- The goal is to choose the coefficients w_j to minimize the sum of the squares of differences (errors) between the predicted and the actual values.
- The sum of the squares of errors:

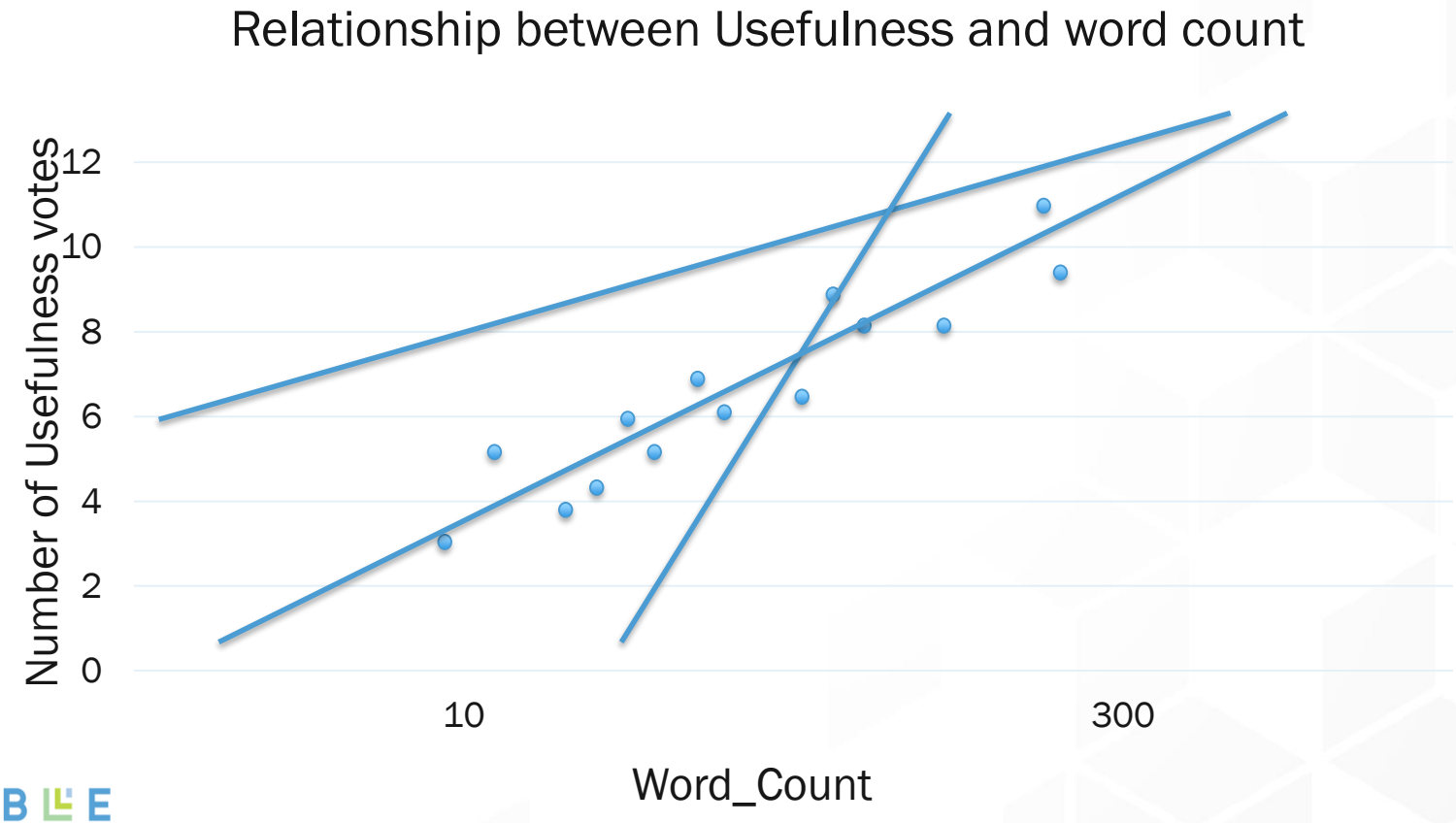
$$\sum_{i=1}^{n}(y^{(i)} - \sum_{j=0}^{k} w_j x_j{}^{(i)})^2$$

  where:
  - x is the attribute values and y is the actual values
  - Superscript denotes that the corresponding instance is *i*th instance
- The proof of the minimization process involves a matrix inversion operation which is over the scope of this boot camp, but readily available as prepackaged software.

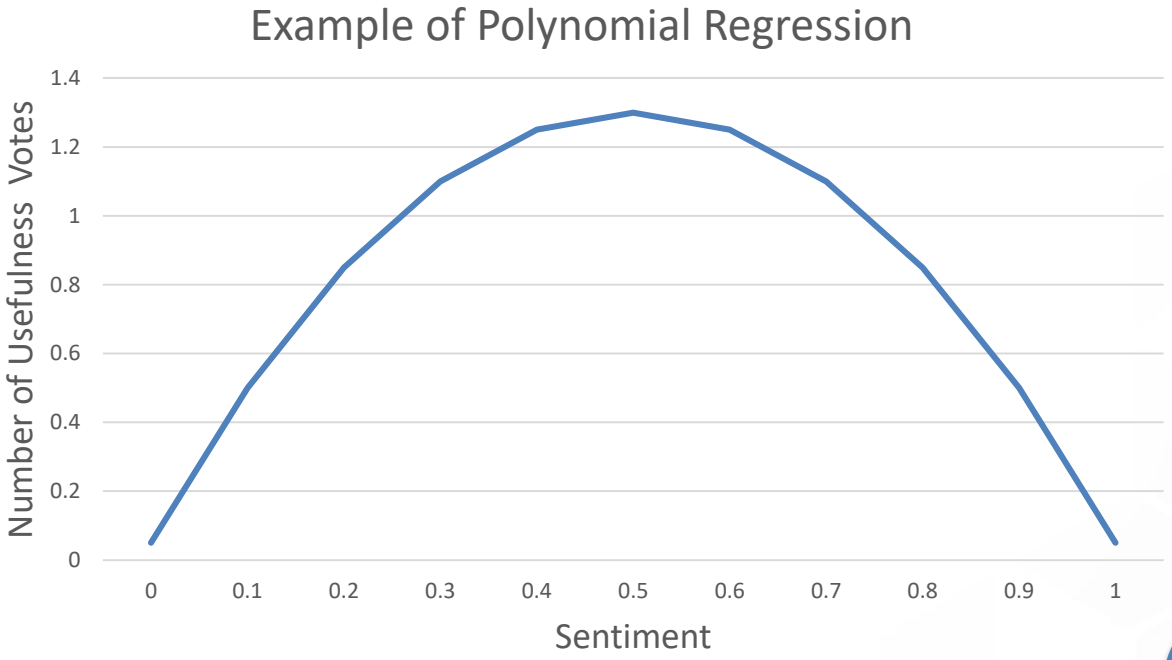
Advantages vs. Disadvantages

- Advantages:
  - Simple, excellent for numeric prediction.
  - Widely used in statistical applications for decades.
- Disadvantages:
  - Assuming linear relationship between dependent and independent variables.
  - Assuming multivariate normality.
  - Assuming no or little multicollinearity.

Common Mistakes in Applying Linear Regression



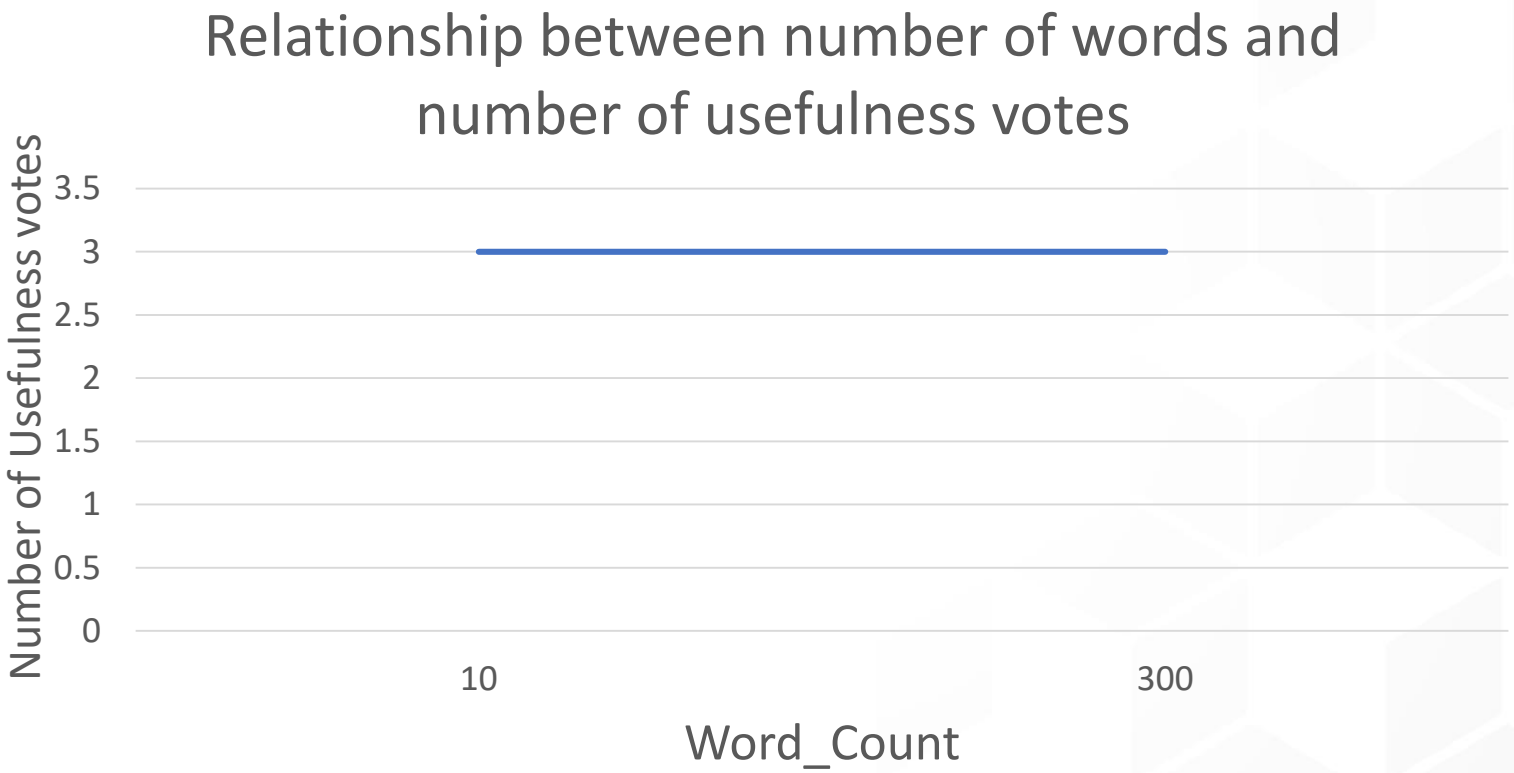
Common Mistakes in Applying Linear Regression: Relationship Between Variables



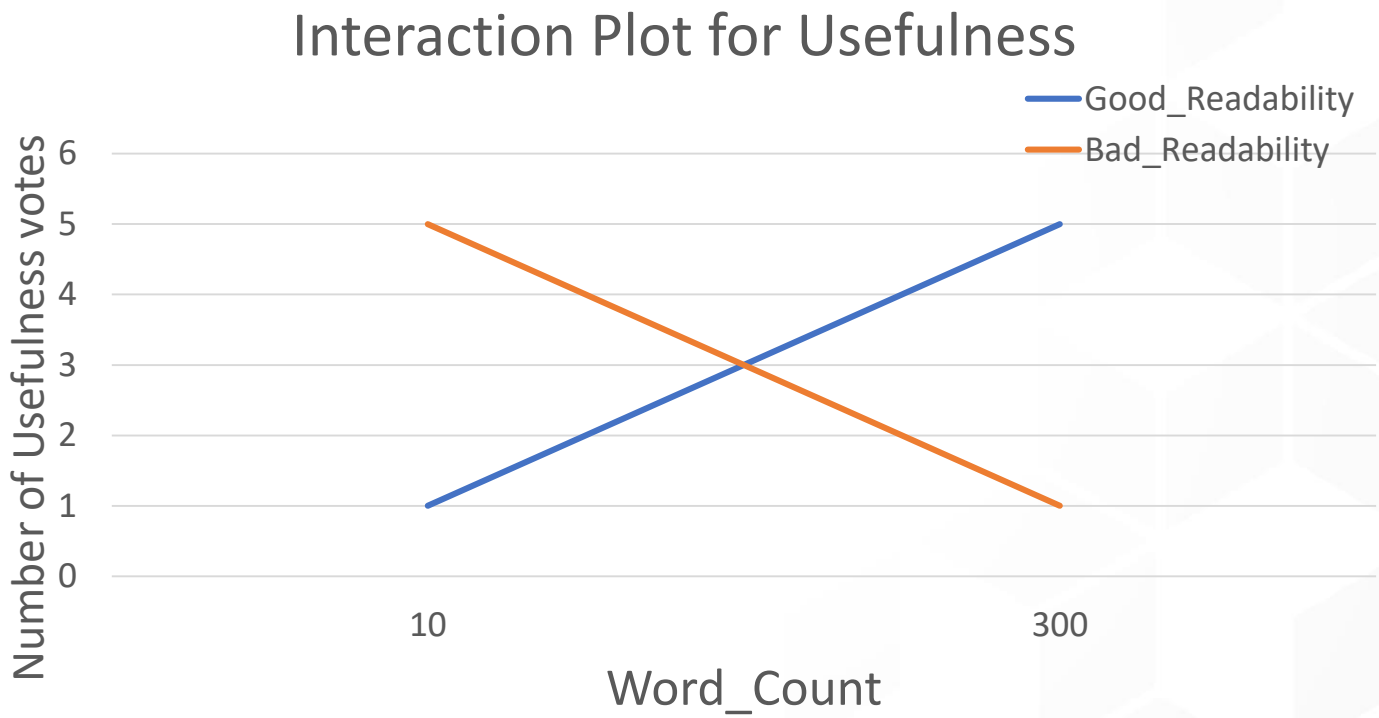
$(x_1 - b_2)^2$

- Polynomial model: $y = a - b_1 * (x_1)$

Common Mistakes in Applying Linear Regression: Interaction of Attributes



Common Mistakes in Applying Linear Regression: Interaction of Attributes (cont ...)



Common Mistakes in Applying Linear Regression: Interaction of Attributes (cont …)

	High readability	Low readability
Large number of words		
Small number of words		

	High readability (0.9)	Low readability (0.1)
Large number of words (0.9)		
Small number of words (0.1)		

* When cell values are multiplication of readability and number of words

Linear Classification: Perceptron Algorithm

$$y = \begin{cases} 1 & \text{if } w_0 + \sum_{j=1}^{n} w_j x_j > 0 \\ 0 & \text{otherwise} \end{cases}$$

parameters learned by the model

predicted value (e.g., 1 = positive, 0 = negative)

Linear Classification: Perceptron Algorithm (cont ...)

$$y = \begin{cases} 1 & \text{if } w_0 + \sum_{j=1}^{n} w_j x_j > 0 \\ 0 & \text{otherwise} \end{cases}$$

test instance

f_1	f_2	f_3
0.5	1	0.2

model weights

w_0	w_1	w_2	w_3
2	-5	2	1

output = 2.0 + (0.50 x -5.0) + (1.0 x 2.0) + (0.2 x 1.0)

output = 1.7

output prediction = positive

Linear Classification: Perceptron Algorithm (cont …)

- What problems can you anticipate with perceptron algorithm?
  - The value ranges from - ∞ to + ∞ which is not ideal for probability.
  - Hard to set the threshold for different classes.

Linear Classifier: Logistic Regression

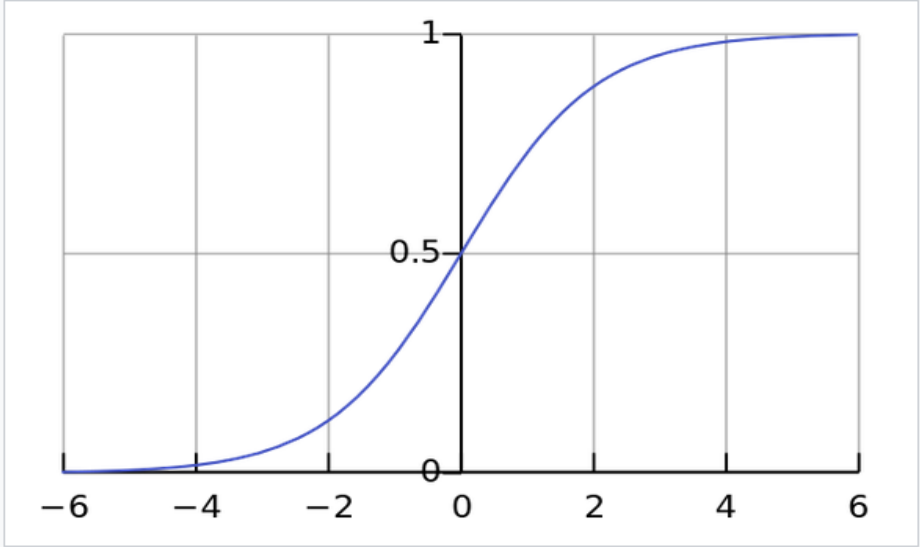


Figure 1. The standard logistic function $\sigma(t)$; note that $\sigma(t) \in (0, 1)$ for all t.

$$\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}$$

when $t = w_0 + \sum_{j=1}^{n} w_j x_j$

(source: https://en.wikipedia.org/wiki/Logistic_regression#/media/File:Logistic-curve.svg)

Any Questions?

Naïve Bayes

Next Class

THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL