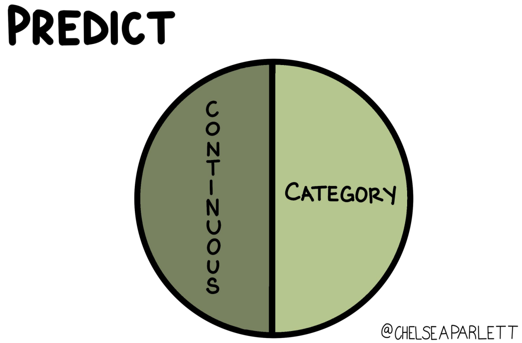


Linear Regression II

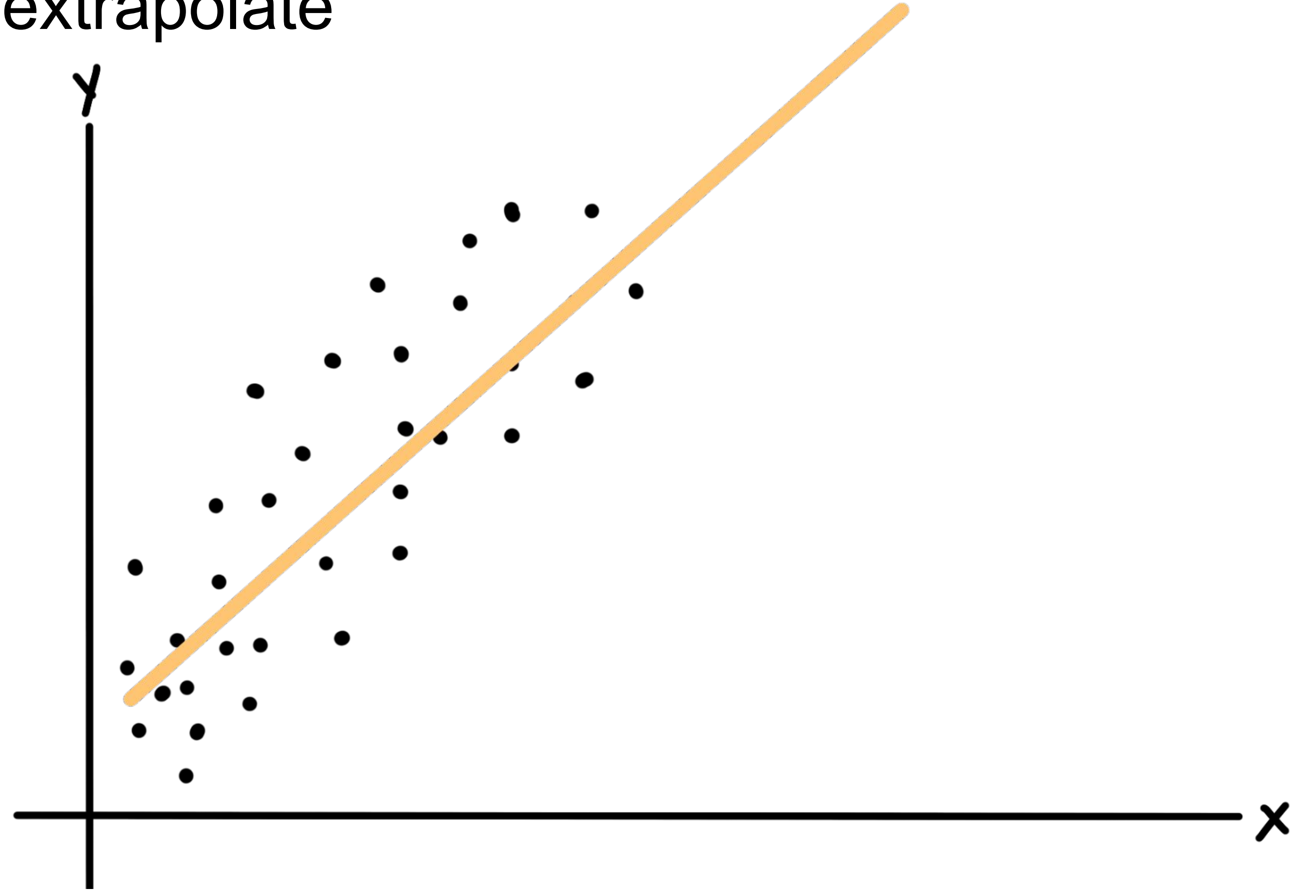
Dr. Chelsea Parlett-Pelleriti

Linear Regression

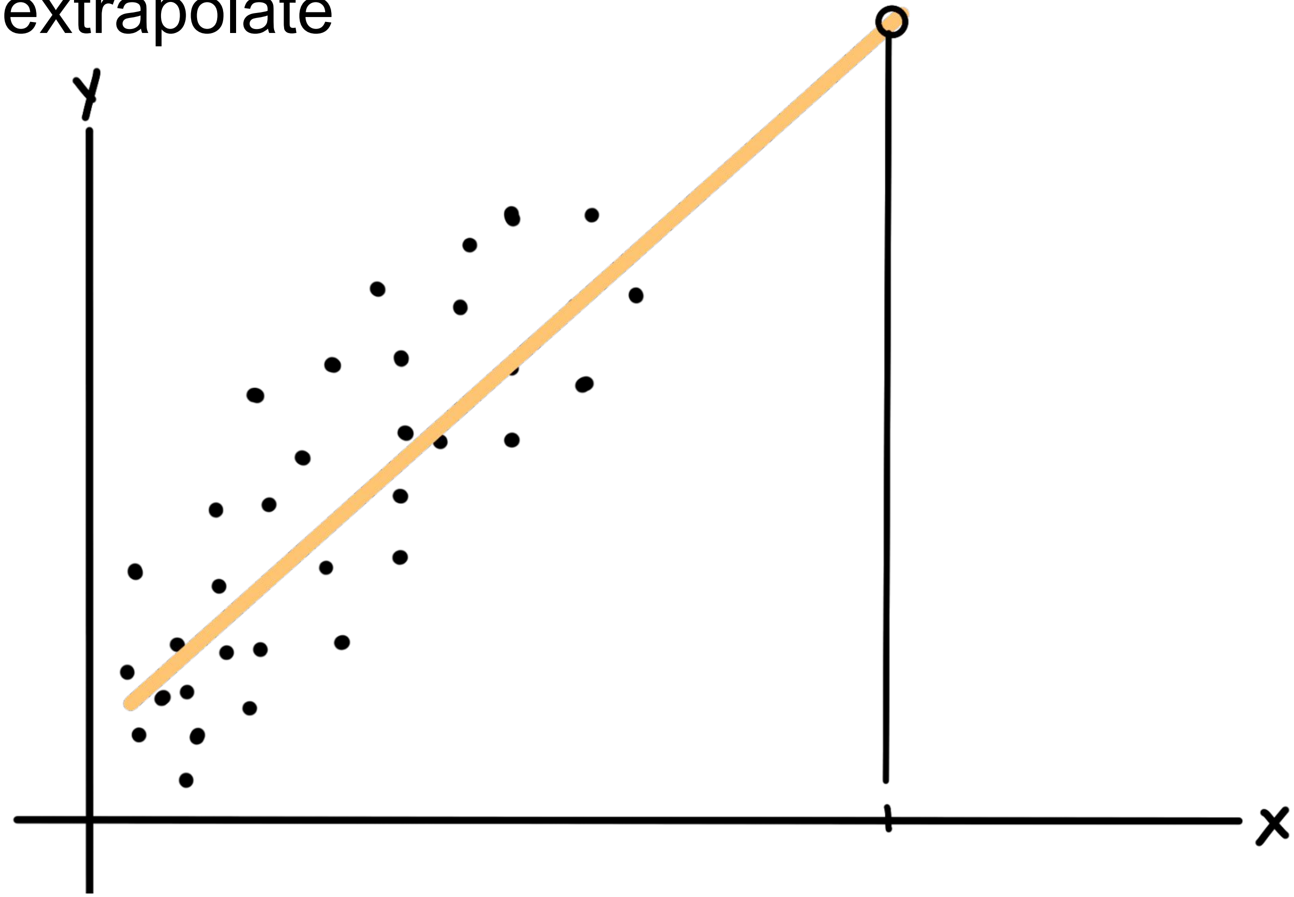
- Extrapolation
- Feature Engineering
- “Non-Linear” Linear Regression
  - Polynomial Regression
  - Interactions
  - Step Functions
  - Basis Functions
  - Regression Splines
  - GAMs

Extrapolation

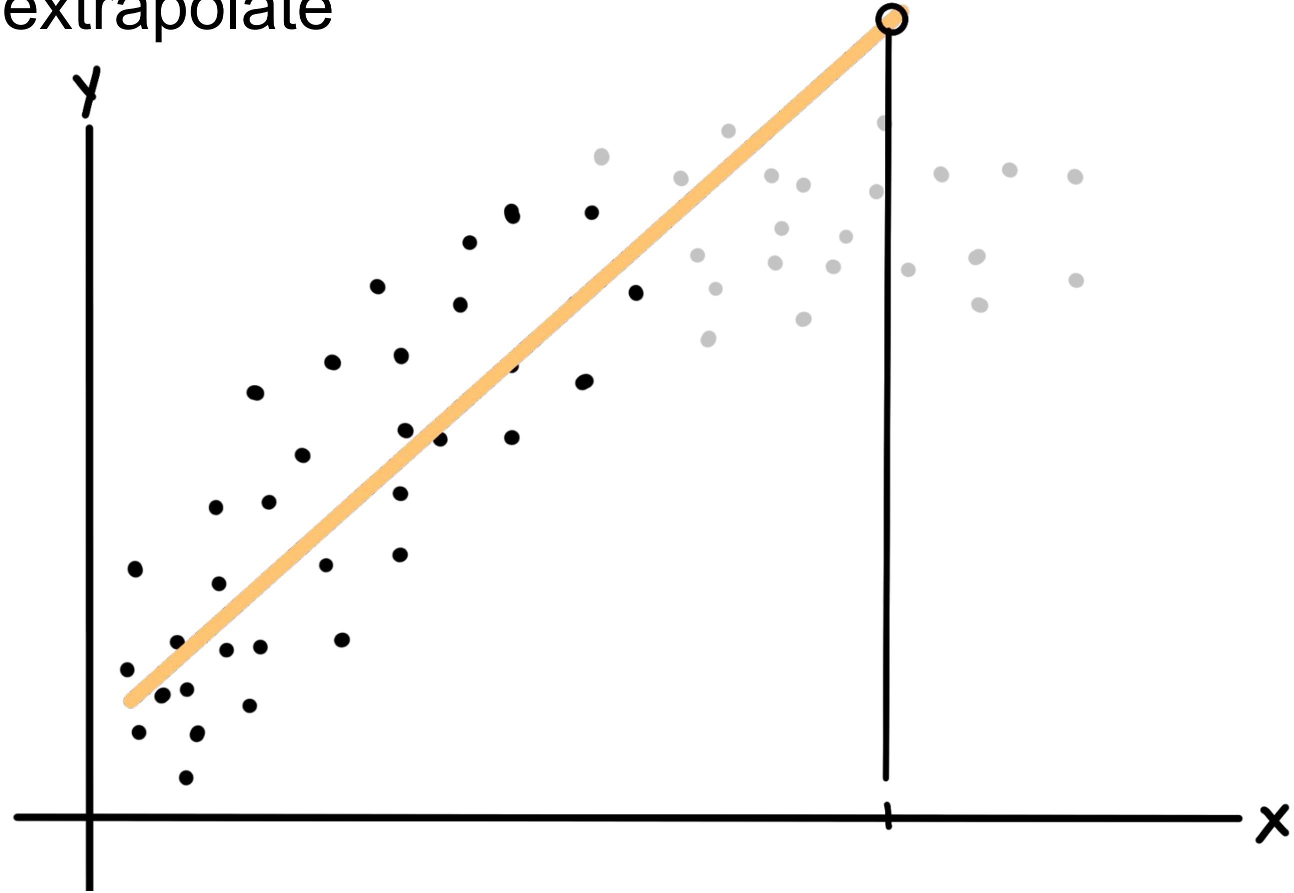
Don’t extrapolate



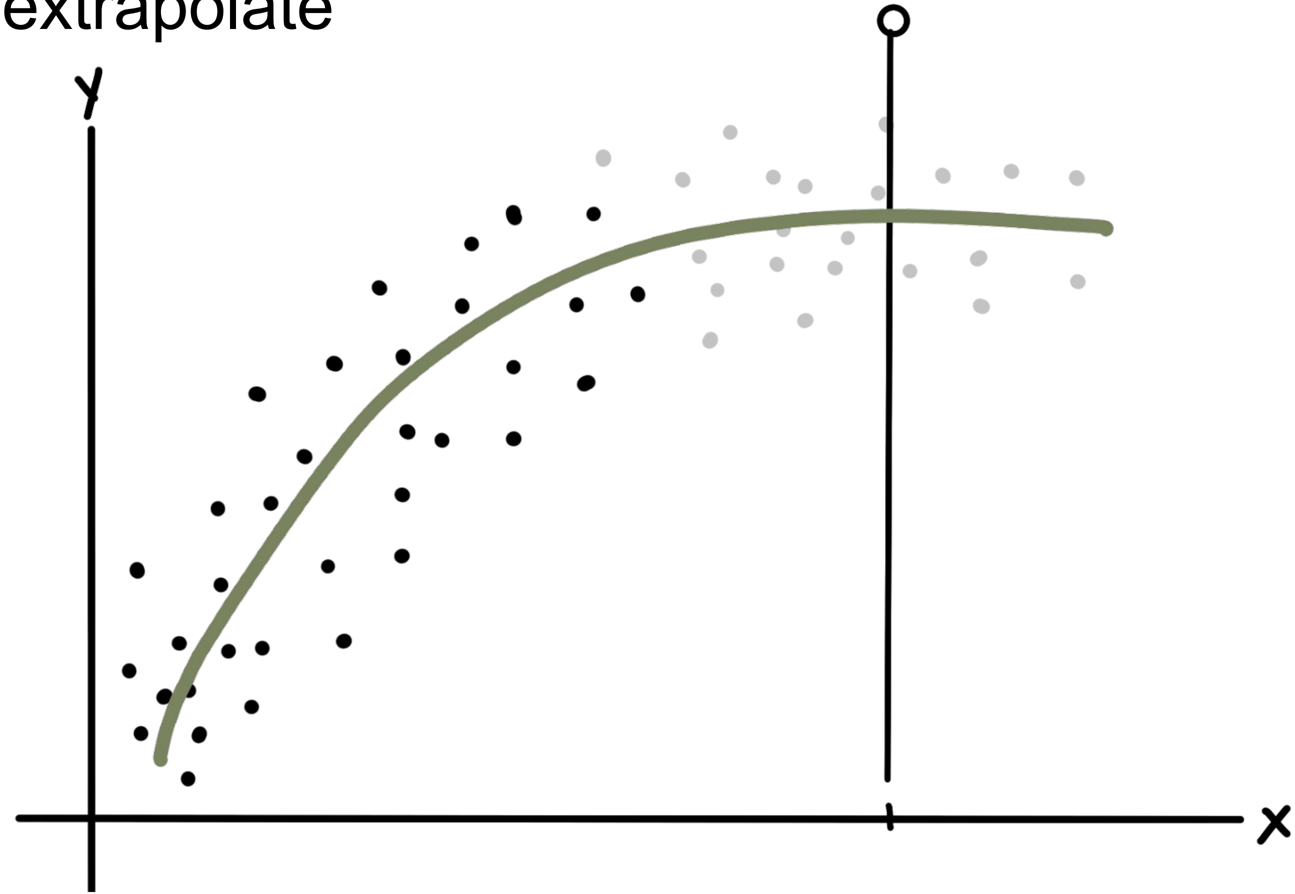
Don’t extrapolate



Don’t extrapolate



Don’t extrapolate



Don’t Extrapolate

Nova Walking Example

Feature Engineering

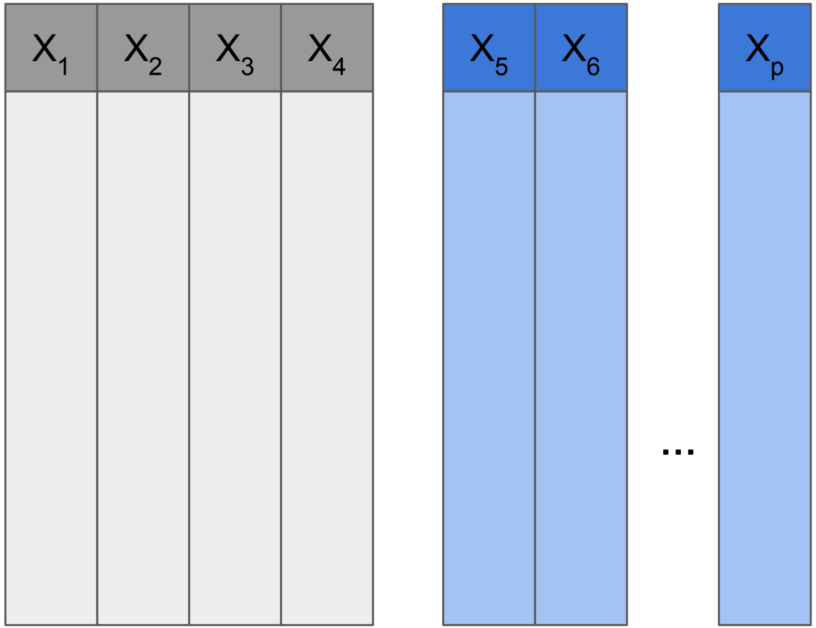
“Feature Engineering”

But what if there IS non-linearity?

X_1	X_2	X_3	X_4

"Feature Engineering"

But what if non-linearity?



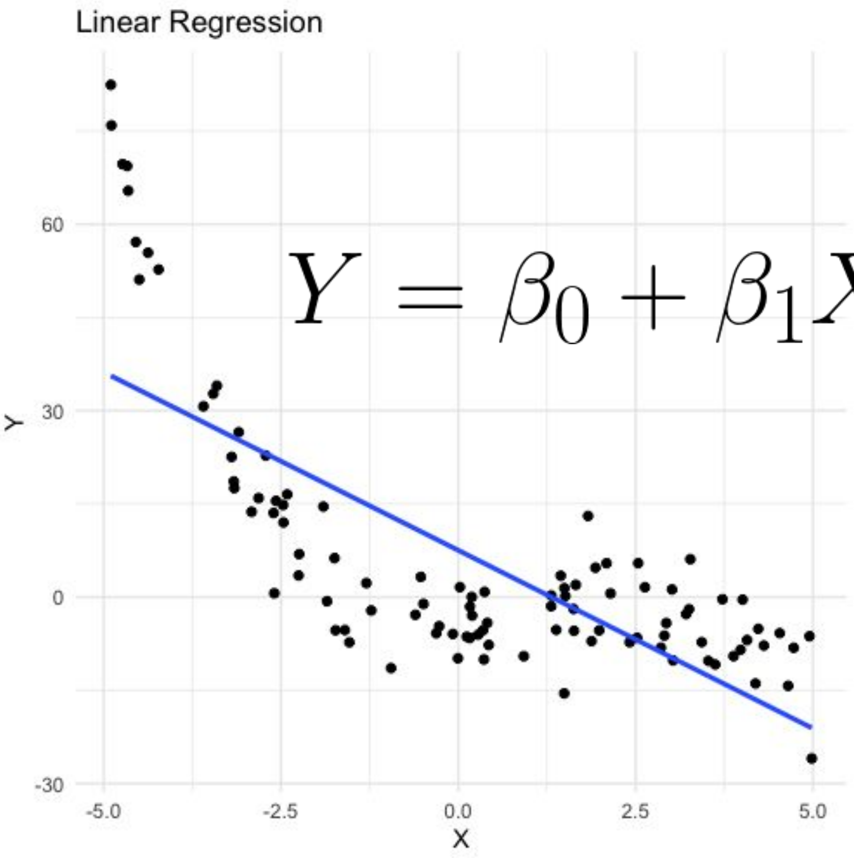
“Non-linear” Linear Regression

Creating New Features

Two Ways to Use “Linear Regression” But Get Non-Linearity

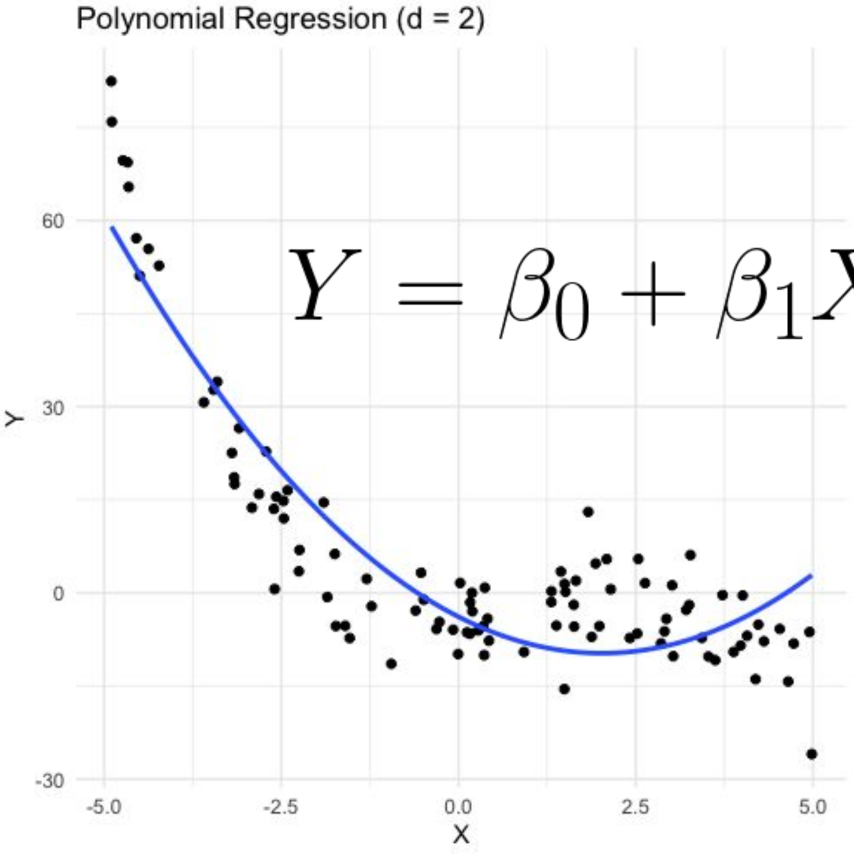
1. Create New Features (Polynomial, GAMs…)⭐
2. Link Functions

Polynomial Regression



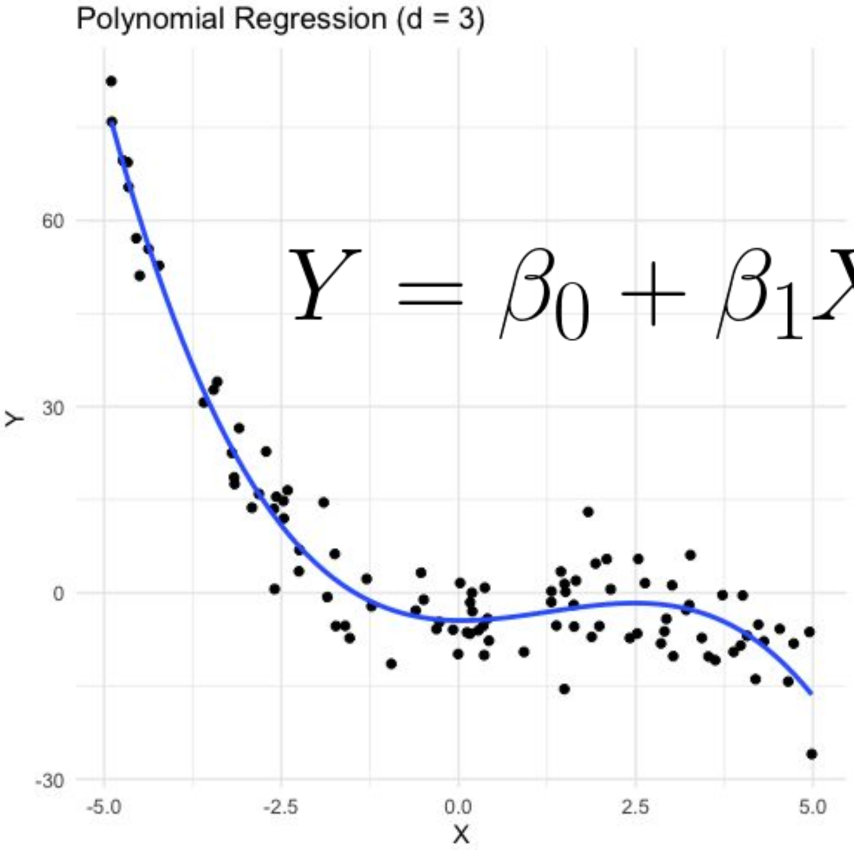
$$Y = \beta_0 + \beta_1 X$$

Polynomial Regression



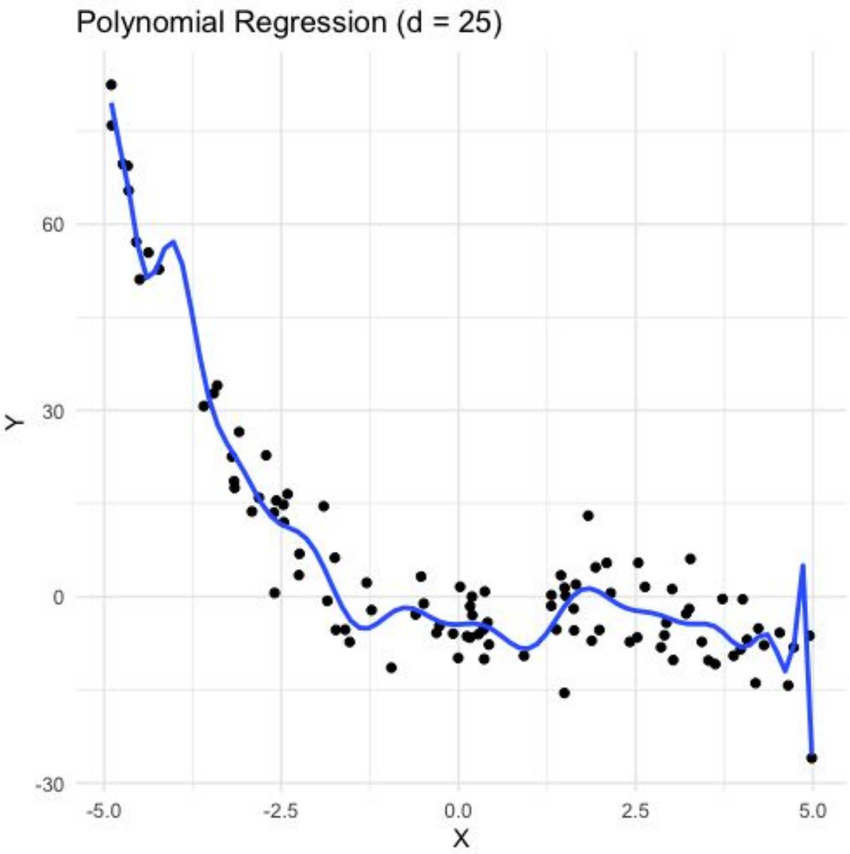
$$Y = \beta_0 + \beta_1 X + \boxed{\beta_2 X^2}$$

Polynomial Regression

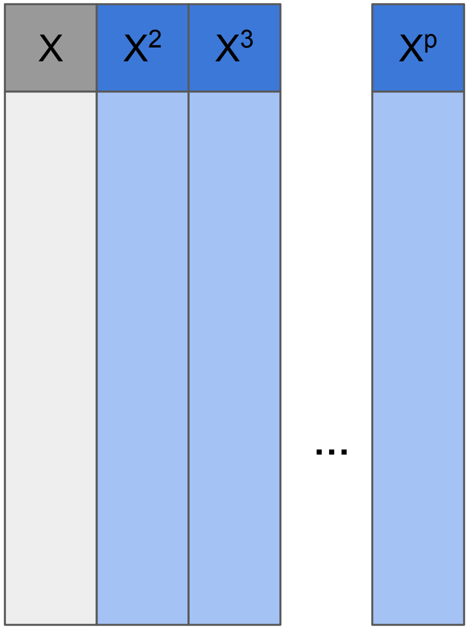


$$Y = \beta_0 + \beta_1 X + \boxed{\beta_2 X^2} + \boxed{\beta_3 X^3}$$

Polynomial Regression



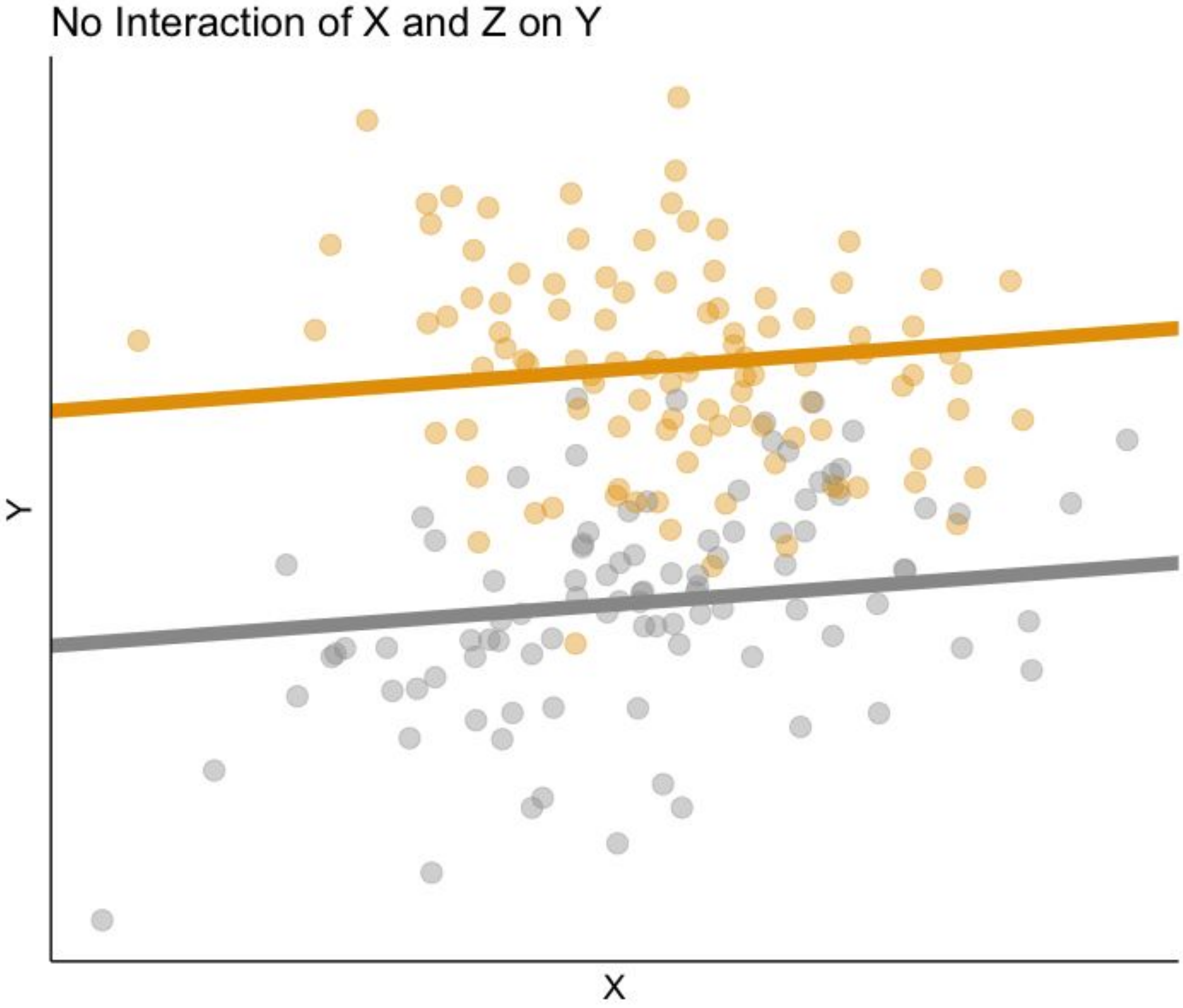
Polynomial Regression



Interactions

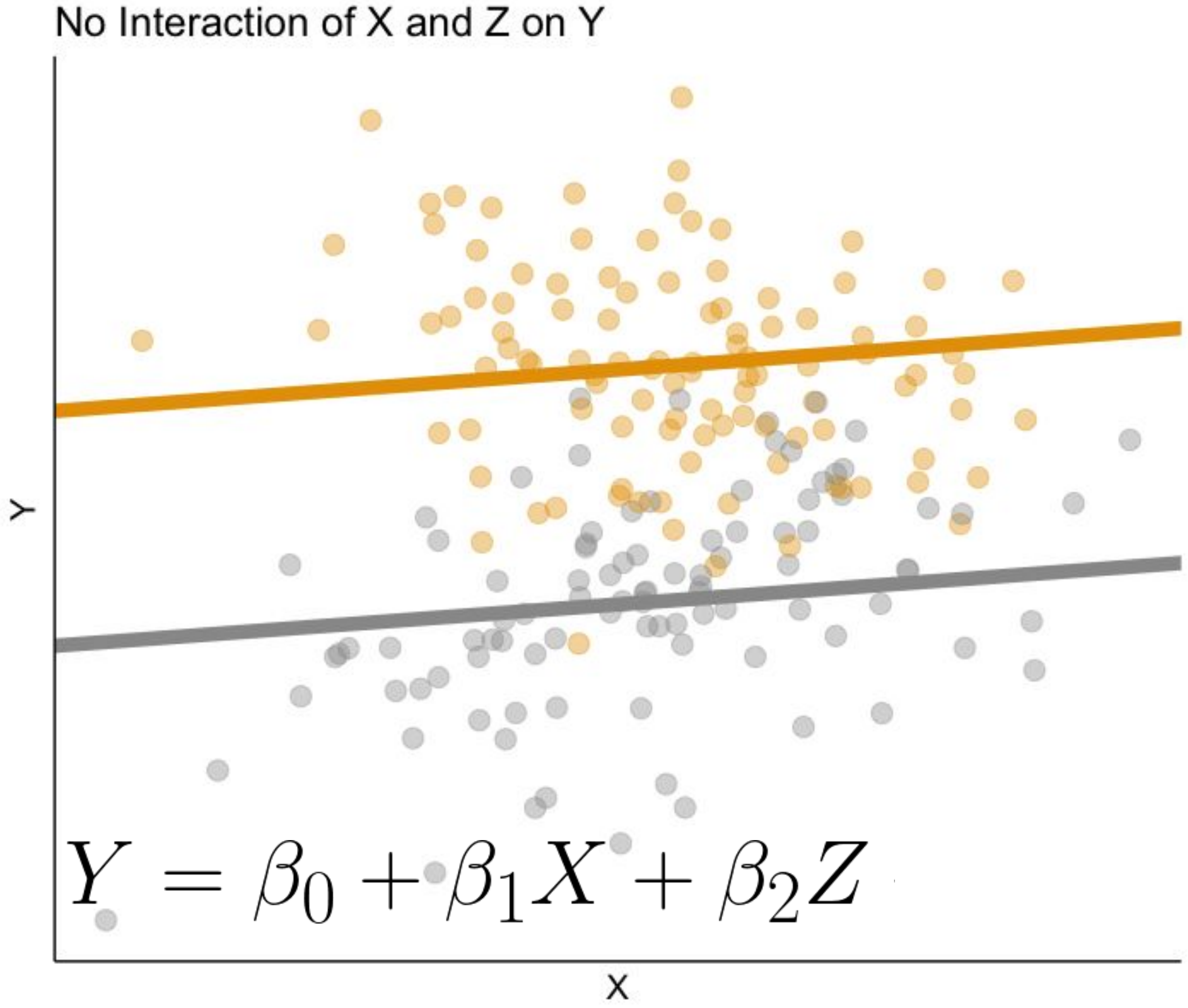
Interactions

Is the relationship between **X** and **Y** different when you consider the value of **Z**?



Interactions

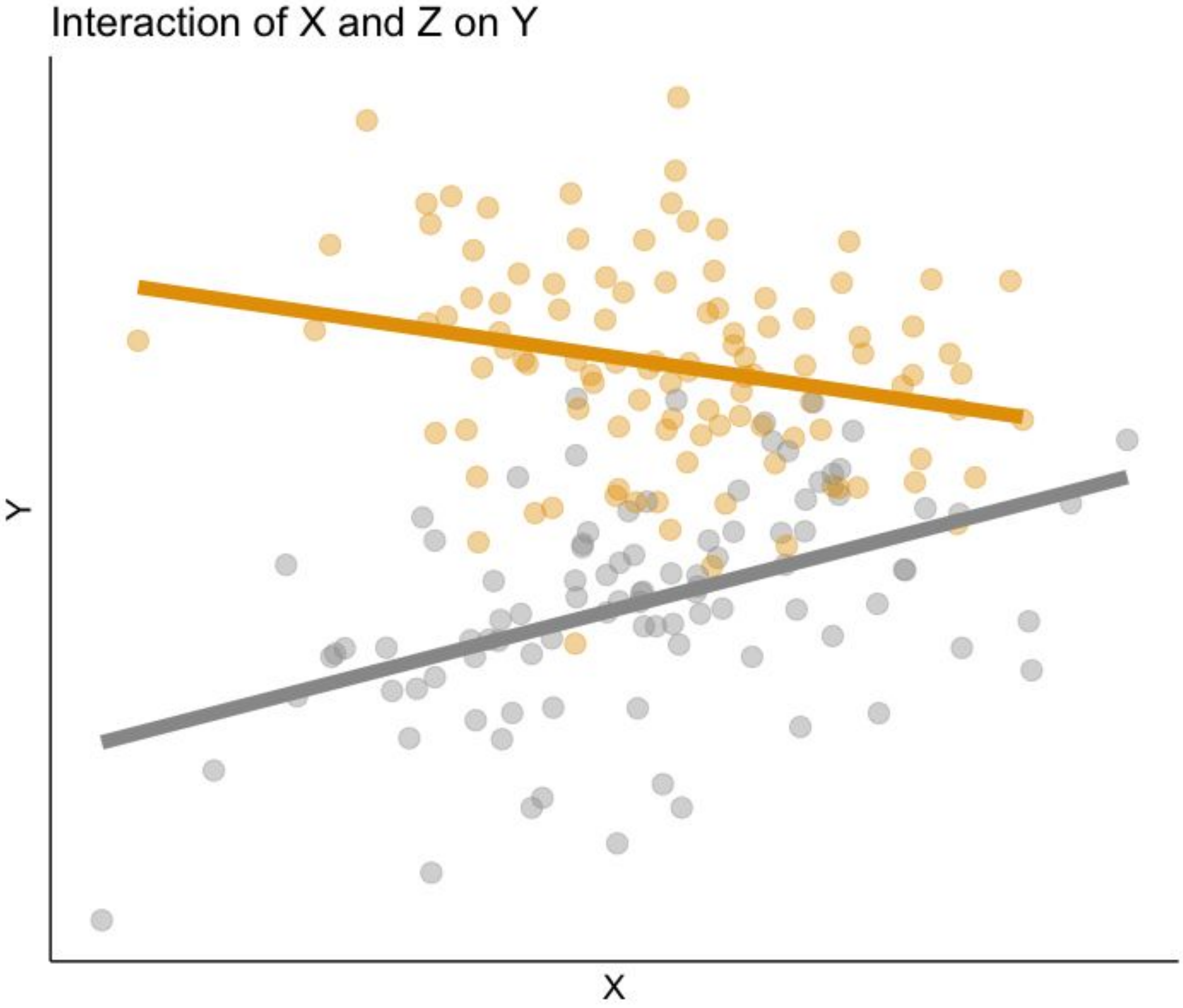
Is the relationship between **X** and **Y** different when you consider the value of **Z**?



$$Y = \beta_0 + \beta_1 X + \beta_2 Z$$

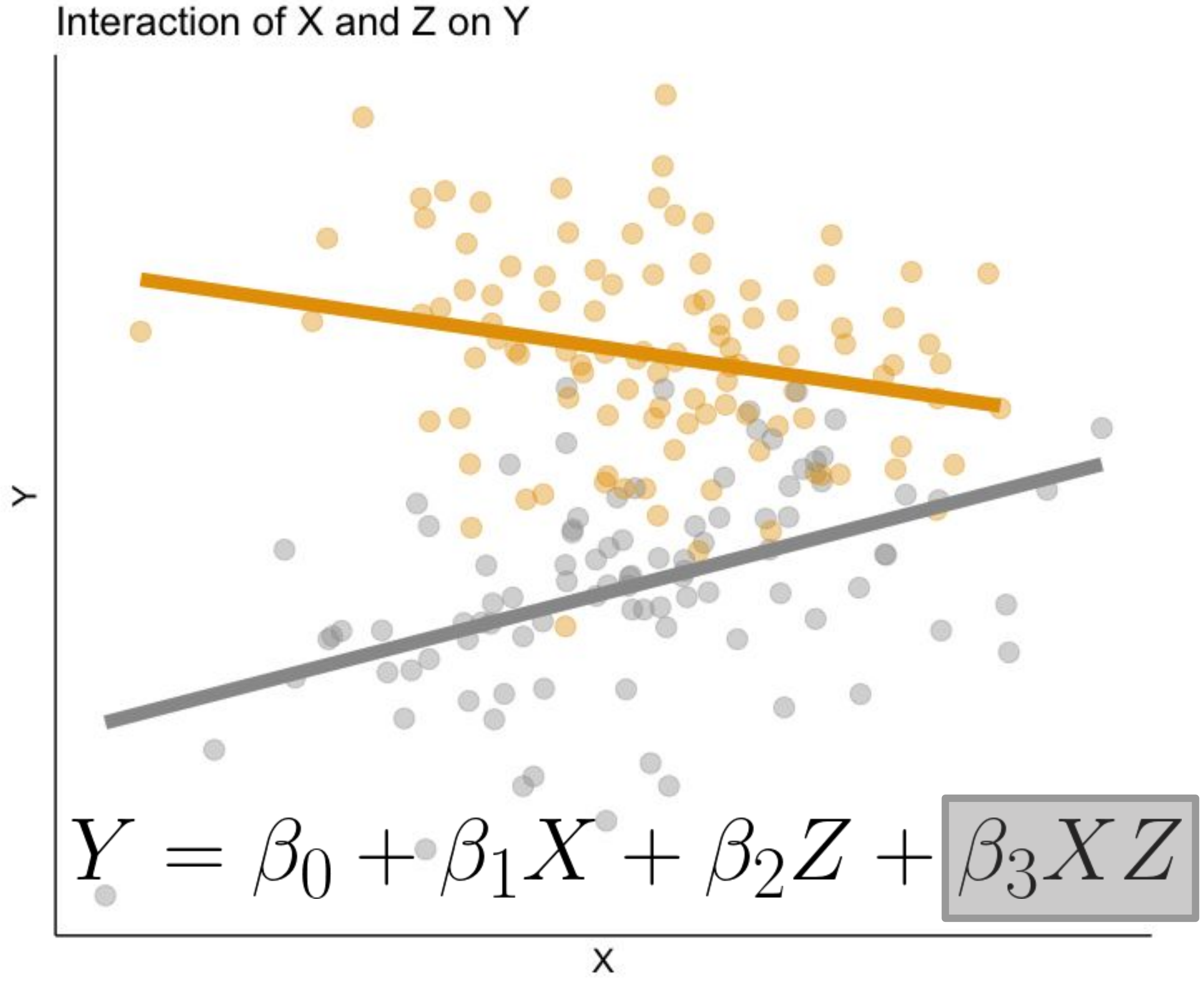
Interactions

Is the relationship between **X** and **Y** different when you consider the value of **Z**?



Interactions

Is the relationship between **X** and **Y** different when you consider the value of **Z**?

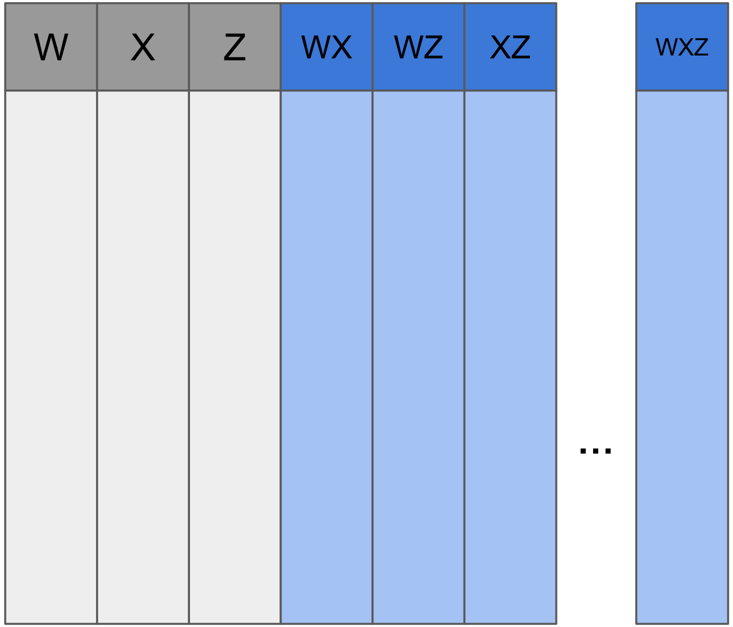


$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \boxed{\beta_3 X Z}$$

Interactions

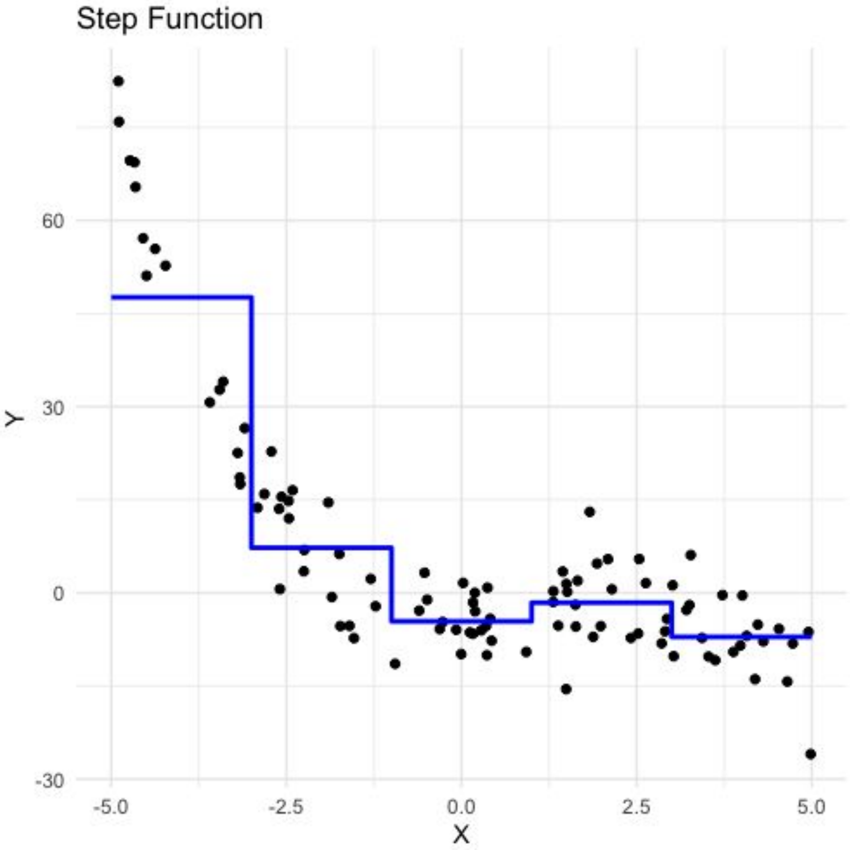
X	Z	XZ

Interactions

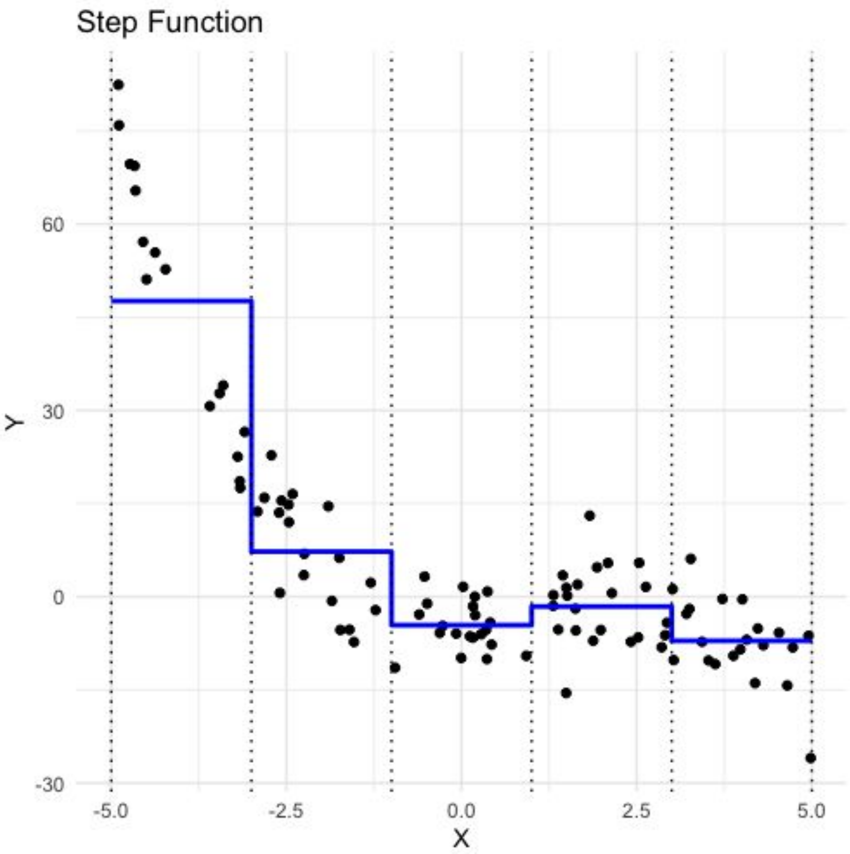


Step Functions

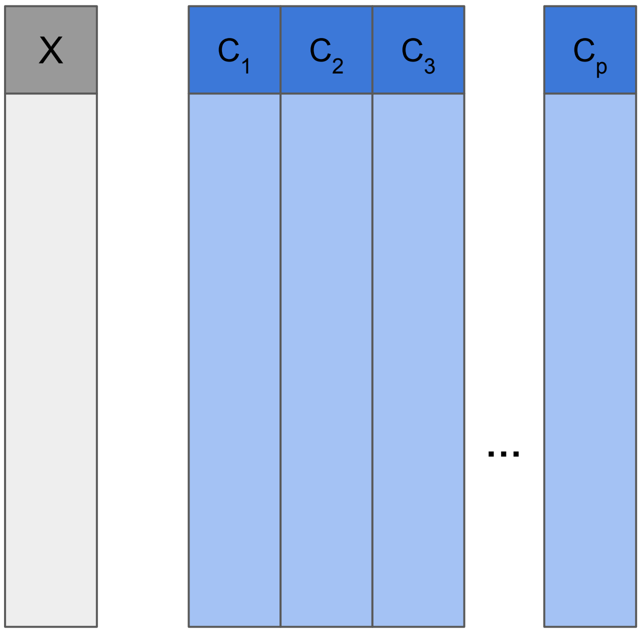
Step Functions



Step Functions



Step Functions



Basis Functions

Basis Functions

$$Y = \beta_0 + \beta_1 f_1(X_1) + \beta_2 f_2(X_2) + \beta_3 f_3(X_3) + \ldots + \beta_p f_p(X_p)$$

Regression Splines

Regression Splines

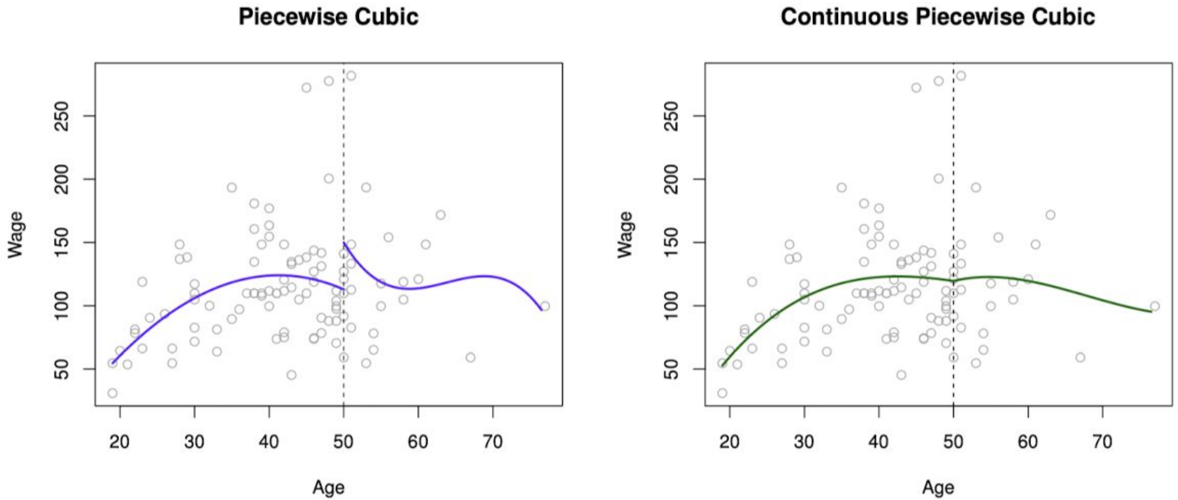
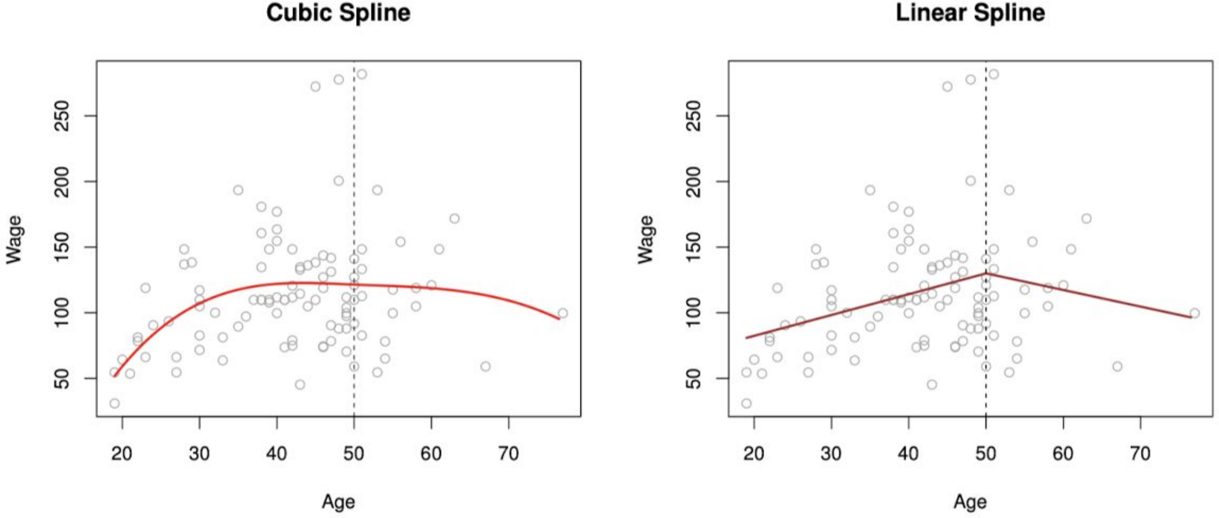


Image from: Introduction to Statistical Learning with applications in Python (Figure 7.3)



Regression Splines

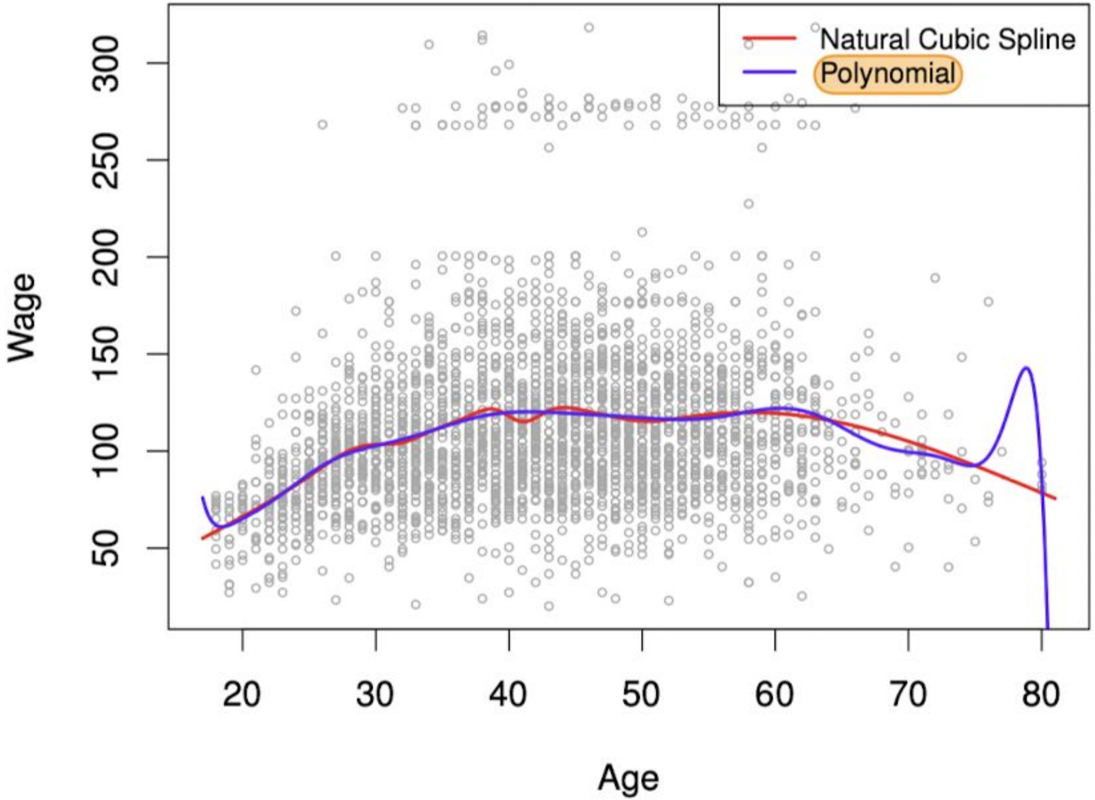


Image from:
Introduction to
Statistical Learning
with applications in
Python (Figure 7.7)

Generalized Additive Models

Generalized Additive Models

$$Y = \beta_0 + \beta_1 f_1(X_1) + \beta_2 f_2(X_2) + \beta_3 f_3(X_3) + ... + \beta_p f_p(X_p)$$

Generalized Additive Models

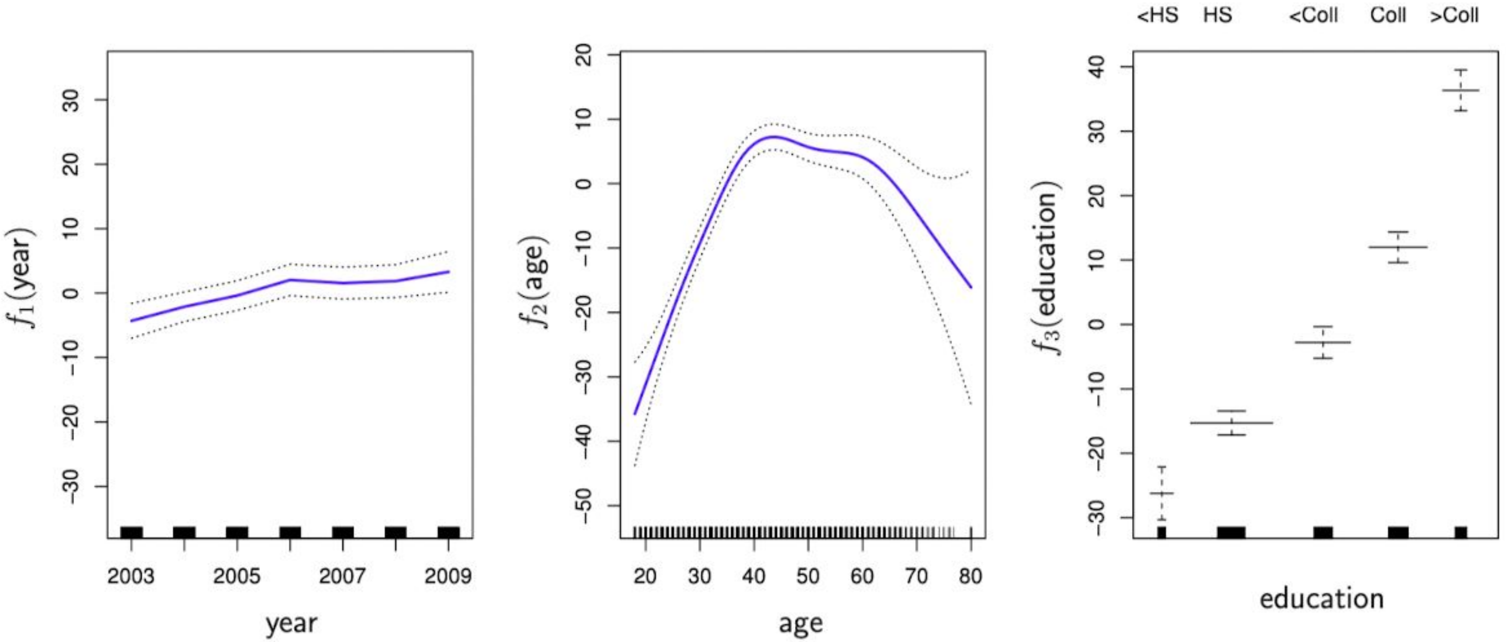


Image from:
Introduction to
Statistical Learning
with applications in
Python (Figure 7.12)