Polynomial Regression and Overfitting

TM Quest

Overview

Overview

What Will we Learn in This Module?

- Polynomial Features
  - Manually creating polynomially features
  - Using Scikit-Learn's built in `PolynomialFeatures` class
  - Bias terms and interaction terms
  - Placing polynomial features into a pipeline
- Overfitting
  - What is overfitting?
  - How to check for overfitting?

Polynomial Regression

What is Polynomial Features?

Creating New Features

We can create new features from old ones and use them in a machine learning model.

Given features X_1 and X_2 we can form polynomials out of them:

- The square feature X_1^2 is every observation of the feature X_1 multiplied by itself.
- The cross-term $X_1 \cdot X_2$ is the product of every observation in X_1 with every observation in X_2.
- We can make infinitely many polynomial features such as $X_1^{17} \cdot X_2^{31}$.

Example Data

Old Features		New Features		
X_1	X_2	X_1^2	X_1X_2	X_2^2
1	4	1	4	16
2	6	4	12	36
7	9	49	63	81
4	2	16	8	4
8	9	64	72	81
4	1	16	4	1
9	0	81	0	0
4	3	16	12	9

Polynomial Regression

Polynomial Regression

Polynomial regression is linear regression where we have made polynomial features.

Instead of approximating the data with a line, we approximate the data with a polynomial curve.

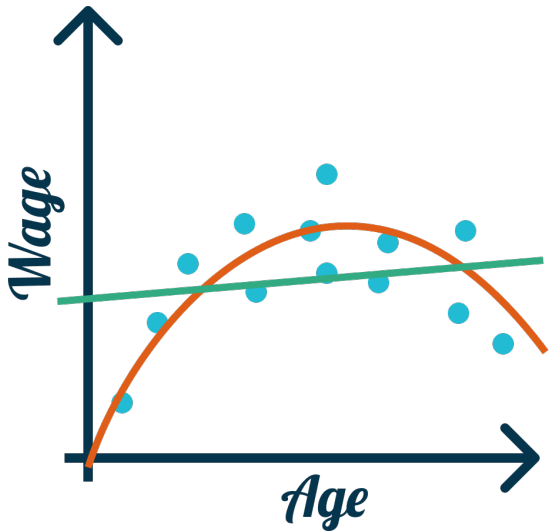
Example

- Feature X representing age.
- Target Y representing wage.

Linear Regression: Find the best straight line $aX + b$ that fits Y.

2nd Order Polynomial Regression: Find the best polynomial $aX^2 + bX + c$ that fits Y.

Straight Line vs. Quadratic Polynomial



Mean Absolute Error

Mean Absolute Error Code

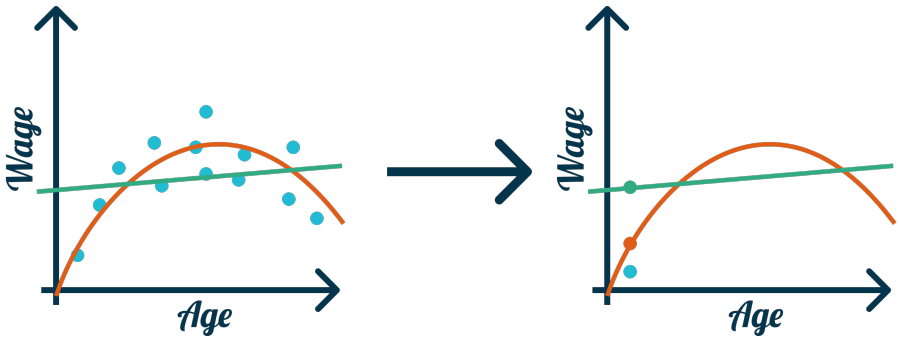
```
import math

# Mean Squared Error
sum((y_pred - y_test) ** 2) / len(y_pred)

# Root Mean Squared Error
math.sqrt(sum((y_pred - y_test) ** 2) / len(y_pred))

# Mean Absolute Error
sum(math.abs(y_pred - y_test)) / len(y_pred)
```

Overfitting & Underfitting

What is Underfitting?

Underfitting

Underfitting is when a model is not complex enough to accurately represent the data.



What is Overfitting?

Overfitting

Overfitting is when a model is so complex that it learns the structure of the training set too well.