

TECHNOLOGIES IN EDUCATION
UNIVERSITY NSU

MICROELECTRONICS
INNOVATIONS
CATALYTIC
MATERIALS
ASSEMBLY
POINT

SCIENTIFIC
LABORATORY
HYBRID
MATERIALS
GEOPHYSICS
ENGINEERING
ENERGY CONSERVATION
BIOTECHNOLOGY
GEOCHEMISTRY
NANOTECHNOLOGY

HIGH
ENERGIES
SEMIOTICS
SCIENCE
MATHEMATICAL MODELING

IT
DEEP
LEARNING
BRAIN
STUDY

COGNITIVE TECHNOLOGIES

DEVELOPMENT
ELEMENTARY
PARTICLES
THE ARCTIC REGIONS
DARK
MATTER

QUANTUM
TECHNOLOGIES
BIOMEDICINE
APPLIED
STUDIES
PHOTONICS
ASTRONOMY
GLOBAL PRIORITY
ASTROPHYSICS
BIOINFORMATICS

LASER
PHYSICS
KNOWLEDGE
ECONOMY

GEOLOGY
ARCHEOLOGY

N* Novosibirsk
State
University
*THE REAL SCIENCE

Non-linear Regression

Luu Minh Sao Khue
Department of Mathematics and Mechanics

Overview

- Non-Linear Regression
- Bias vs Variance
- Ridge Regression
- Lasso Regression
- Data Visualization

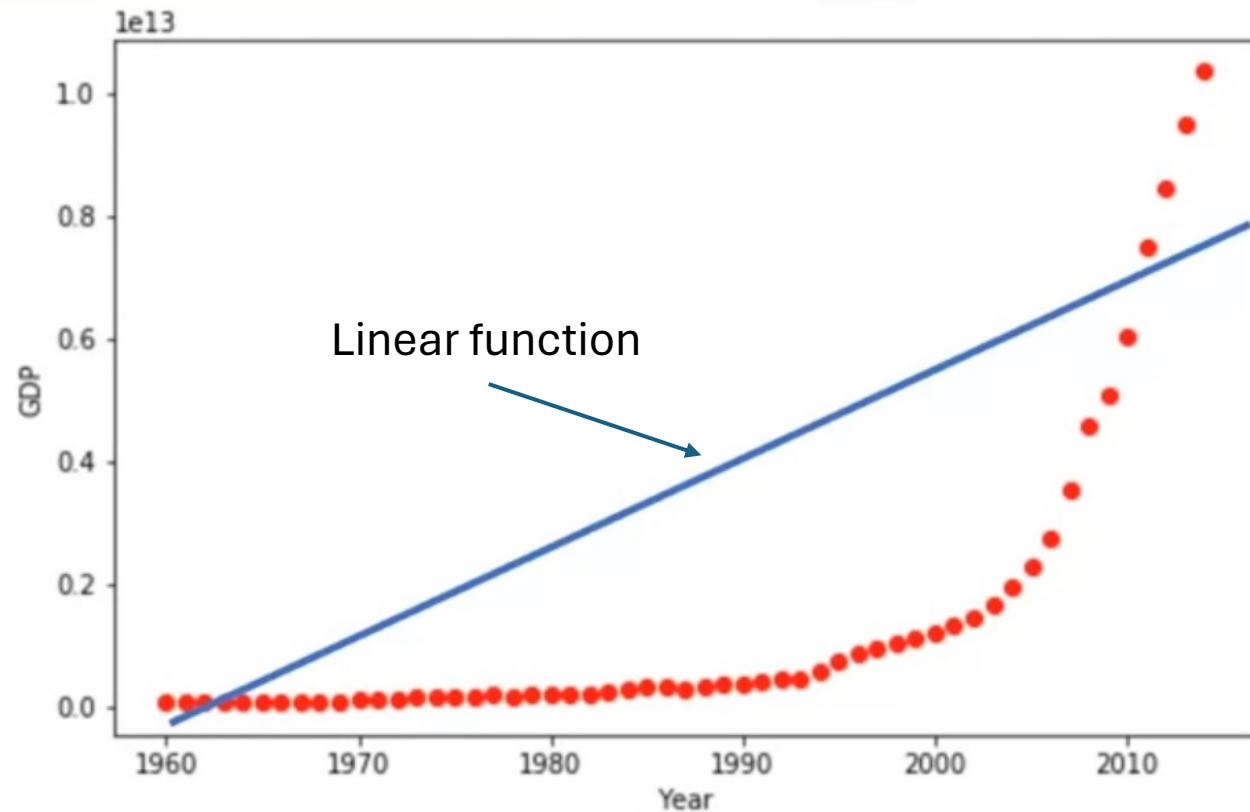
Non-Linear Regression

Non-Linear Regression

- **Non-linear regression** is a type of regression analysis in machine learning where the relationship between the independent variable(s) and the dependent variable cannot be modeled by a straight line (linear equation). Instead, it involves a **non-linear function** of the parameters to model more complex relationships.

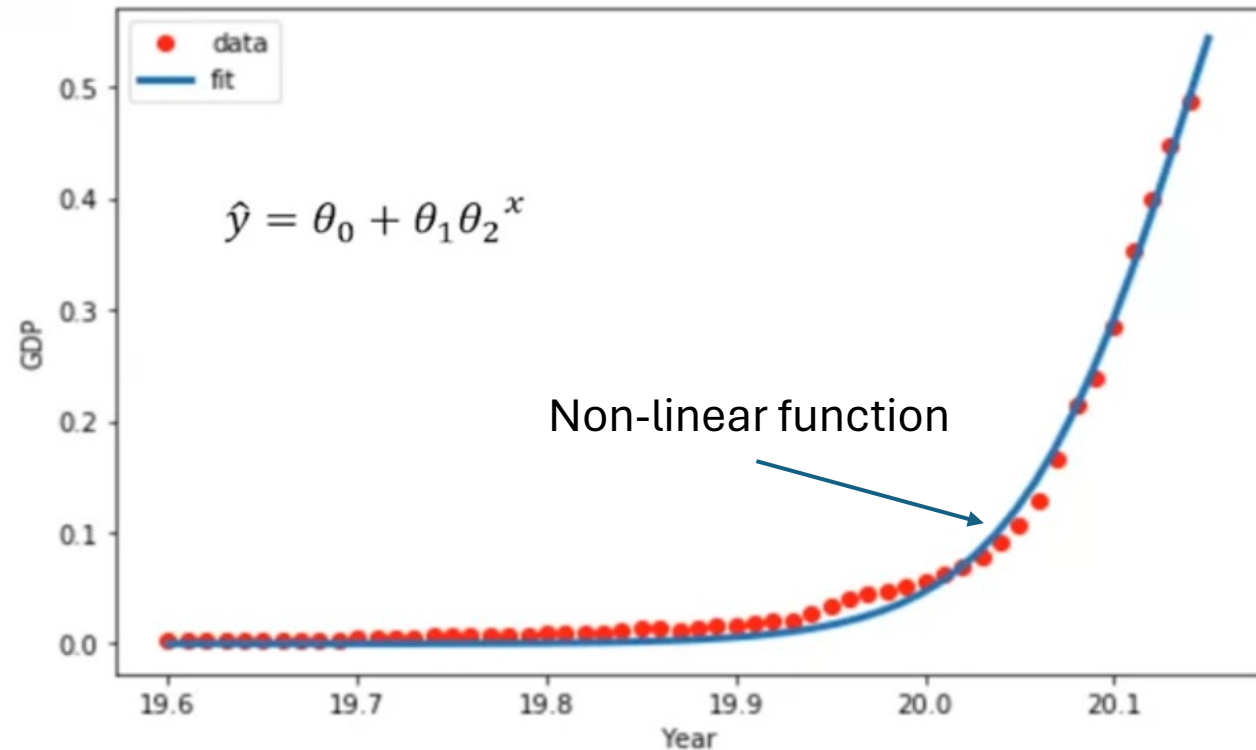
Non-Linear Regression

	Year	Value
0	1960	5.918412e+10
1	1961	4.955705e+10
2	1962	4.668518e+10
3	1963	5.009730e+10
4	1964	5.906225e+10
5	1965	6.970915e+10
6	1966	7.587943e+10
7	1967	7.205703e+10
8	1968	6.999350e+10
9	1969	7.871882e+10
...



Non-Linear Regression

	Year	Value
0	1960	5.918412e+10
1	1961	4.955705e+10
2	1962	4.668518e+10
3	1963	5.009730e+10
4	1964	5.906225e+10
5	1965	6.970915e+10
6	1966	7.587943e+10
7	1967	7.205703e+10
8	1968	6.999350e+10
9	1969	7.871882e+10
...



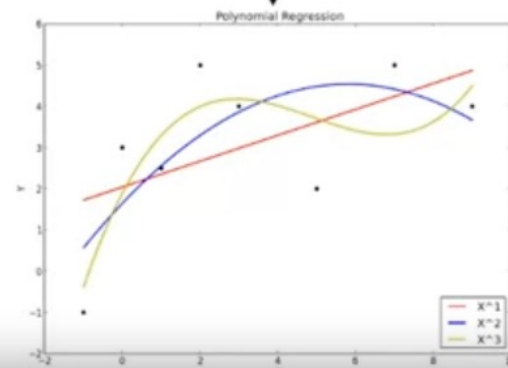
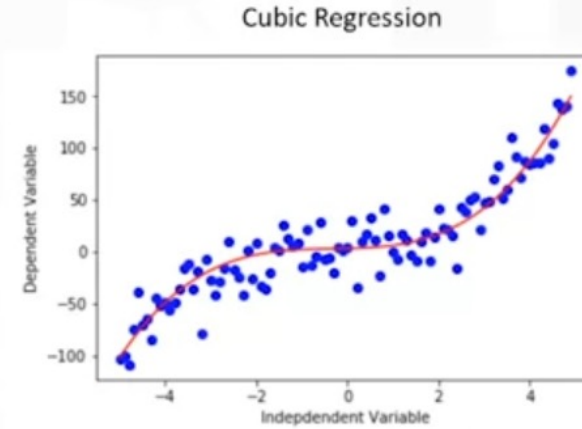
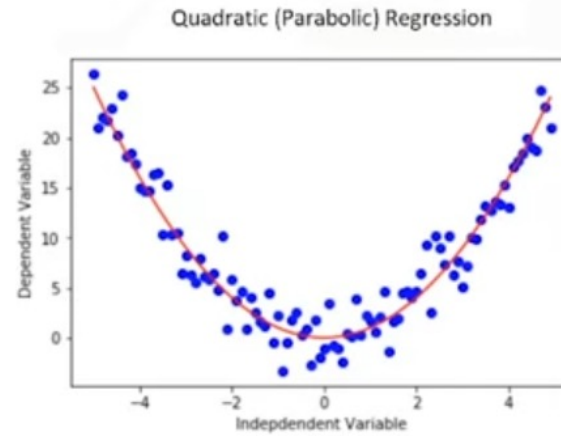
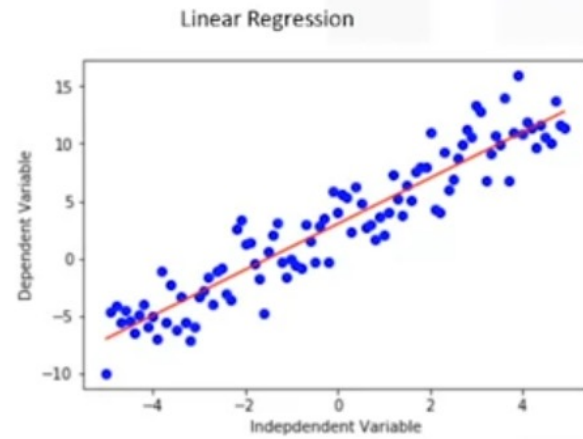
How to identify non-linear relationship

x	y	$\frac{\text{Difference in } y}{\text{Difference in } x}$
3	15	
+2	+8	$\frac{8}{2} = 4$
5	23	
+4	+10	$\frac{10}{4} = 2.5$
9	33	
+2	+8	$\frac{8}{2} = 4$
11	41	
+2	+12	$\frac{12}{2} = 6$
13	43	

All are not same

Visualize

Polynomial Regression



Limitations of Polynomial Regression

- **Overfitting:**
 - Increasing the degree of a polynomial improves the fit on training data but worsens performance on new data.
 - The model becomes too specific to the training set and captures noise.
- **Unpredictable Behavior:**
 - High-degree polynomials diverge rapidly outside the training data range:
 - **Odd degrees:** Shoot off in opposite directions.
 - **Even degrees:** Shoot off in the same direction.
- **When to Use Polynomial Regression:**
 - Only when the underlying relationship is **known to follow a polynomial pattern** (e.g., natural laws like projectile motion).

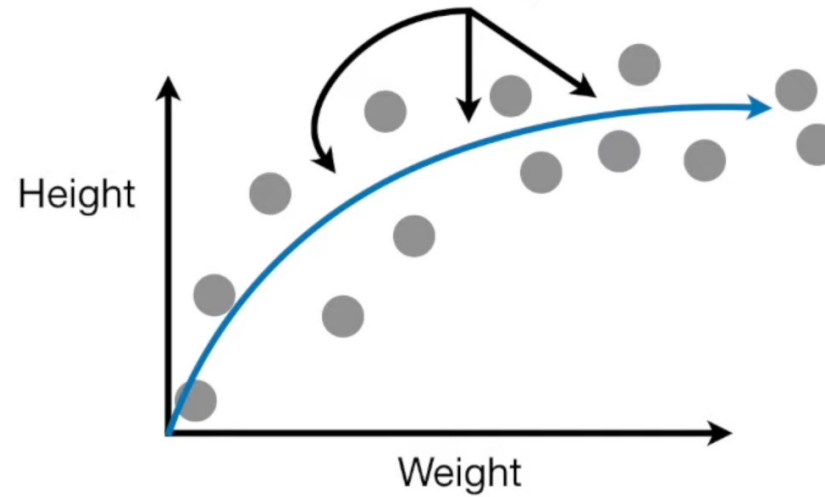
Linear vs Non-Linear Regression

	Linear Regression	Non-Linear Regression
Model Form	Linear equation	Non-linear equation
Complexity	Simple, easy to interpret	Flexible, hard to interpret
Parameters Estimation	Closed-form solution	Iterative optimization
Computation	Low	High

Bias vs Variance

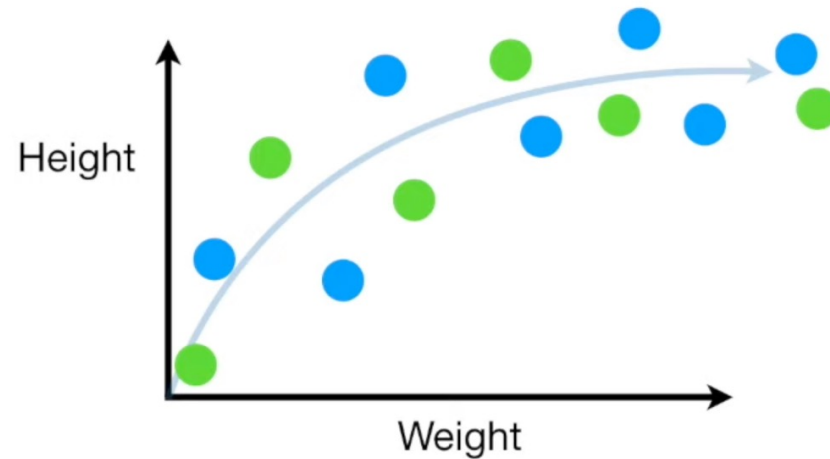
Bias vs Variance

...but, in this case, we don't know the formula, so we're going to use two machine learning methods to approximate this relationship.

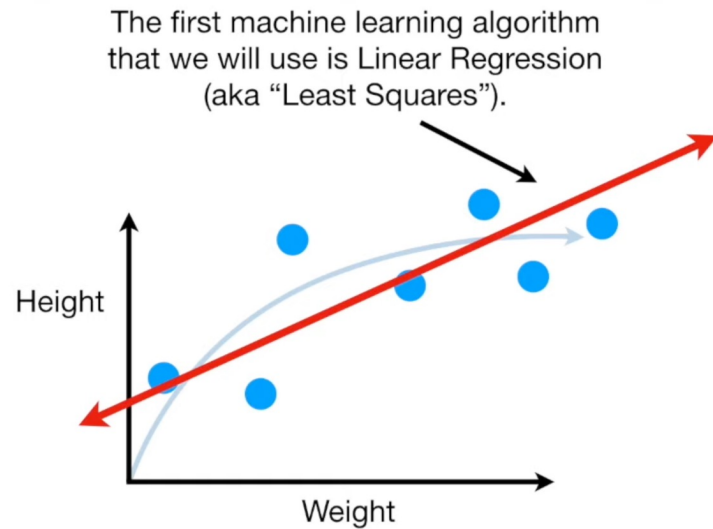


Bias vs Variance

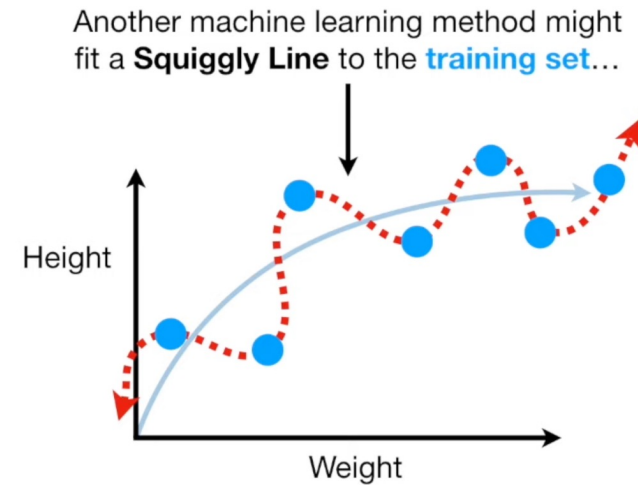
The first thing we do is split the data into two sets, one for training the machine learning algorithms and one for testing them.



Bias vs Variance



High Bias

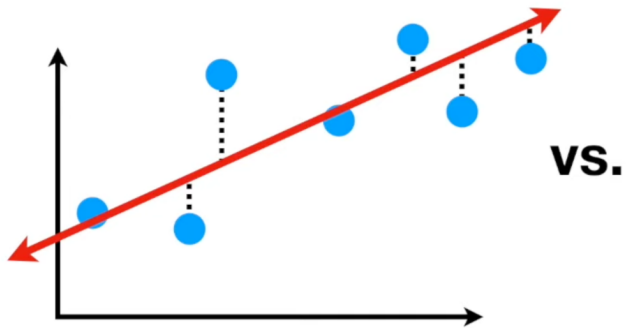


High Variance

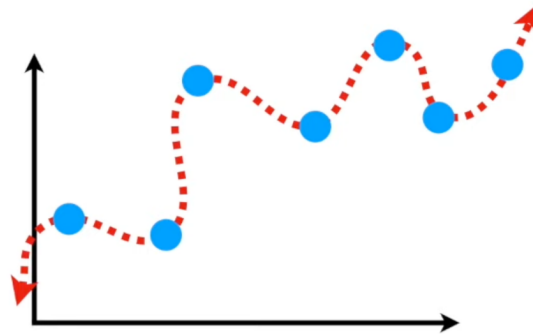
Source: <https://www.youtube.com/watch?v=EuBBz3bl-aA>

Bias vs Variance

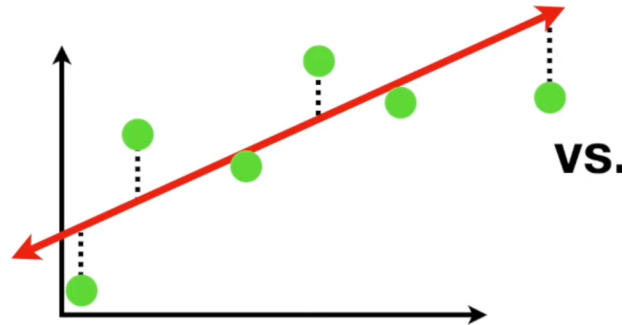
In the contest to see whether the **Straight Line** fits the **training set** better than the **Squiggly Line**...



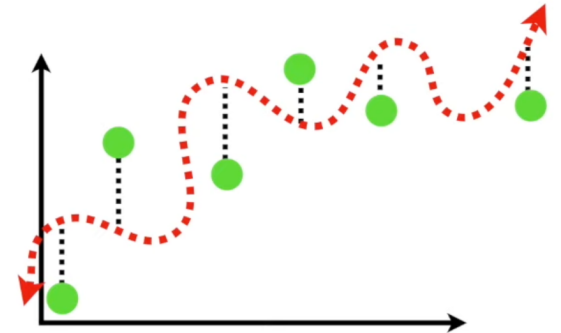
VS.



In the contest to see whether the **Straight Line** fits the **testing set** better than the **Squiggly Line**...



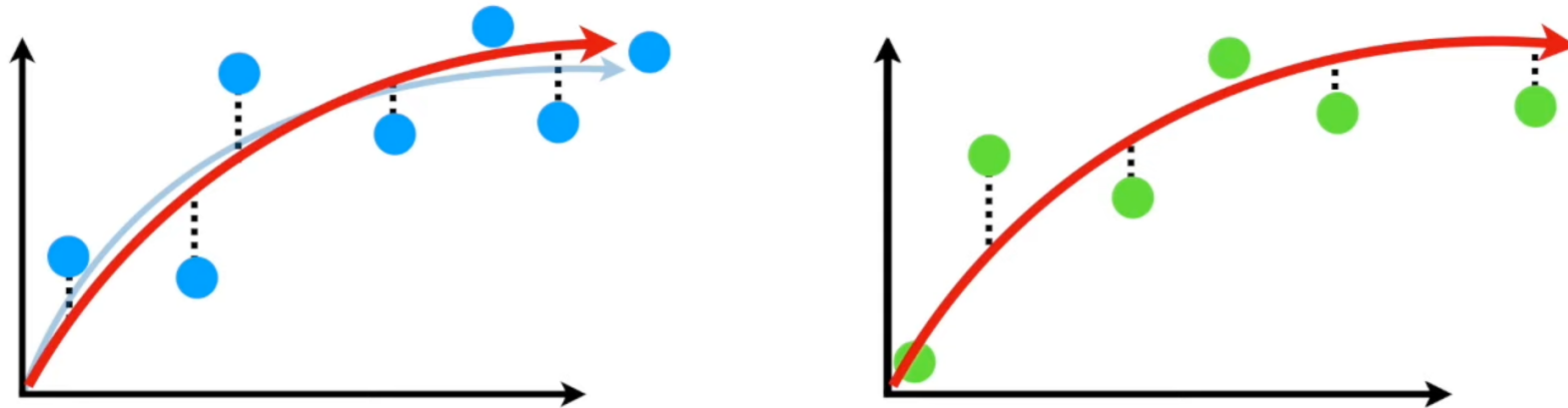
VS.



Source: <https://www.youtube.com/watch?v=EuBBz3bl-aA>

Bias vs Variance

Low Bias Low Variance



Bias vs Variance

- Regularization -> add penalties to the loss function
- Bagging -> reduce variance
- Boosting -> reduce bias

Ridge Regression

Regularization

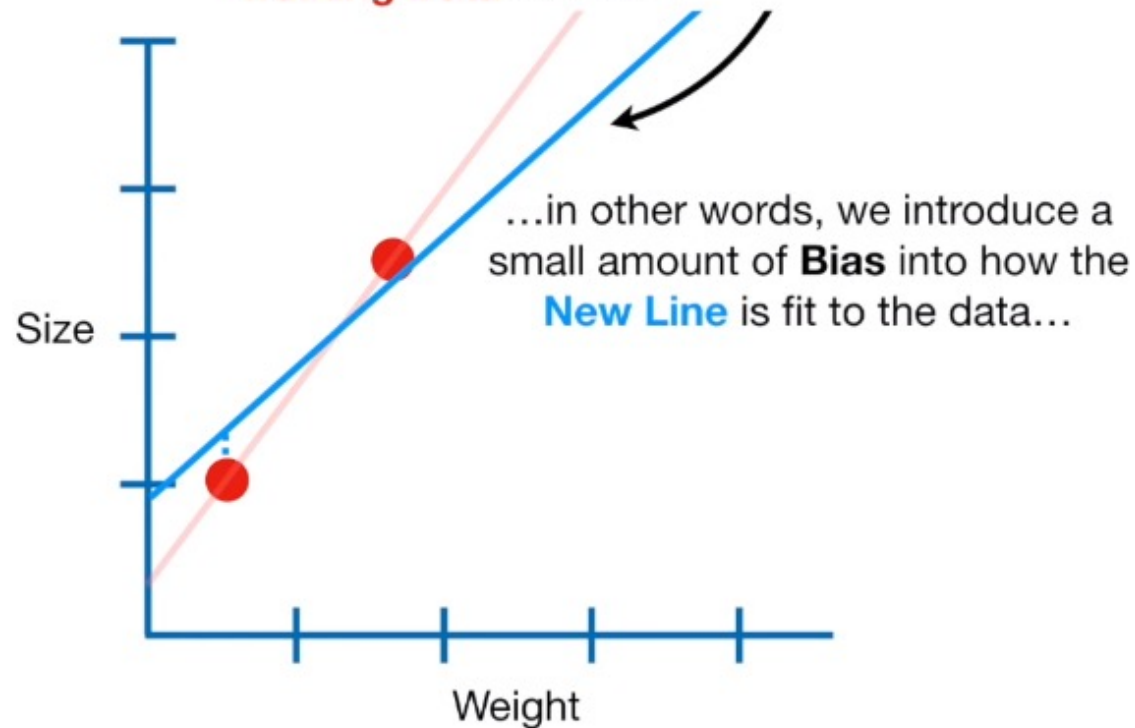
- Regularization is a technique used in machine learning and statistics to improve the performance and generalization of a model by adding a penalty to the loss function. This penalty discourages the model from becoming too complex or overly reliant on specific features, which can lead to **overfitting**.

Ridge Regression

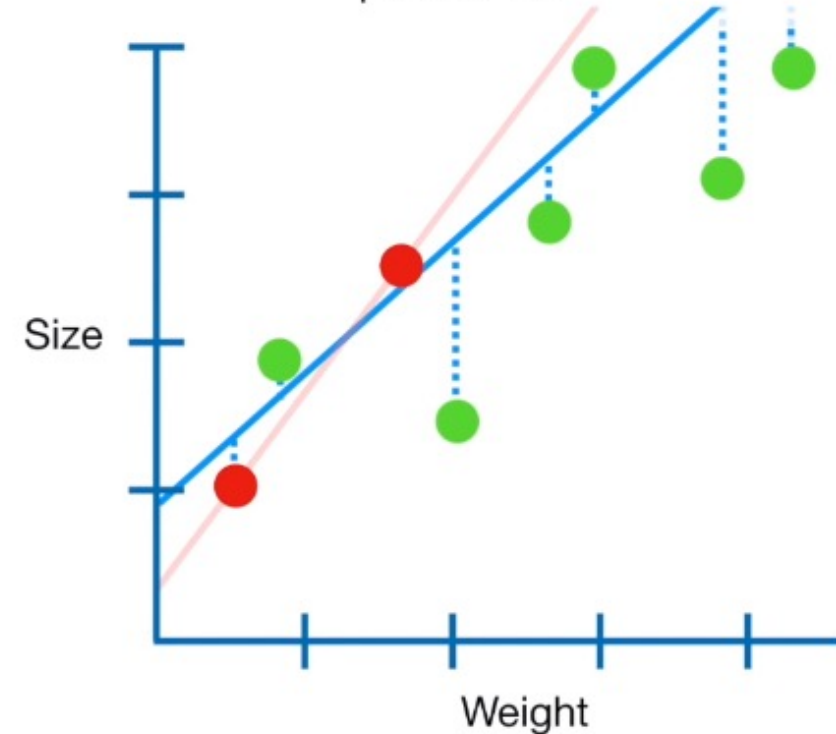
- **Ridge regression**, also known as **L2 Regularization**, is a linear regression technique that introduces a penalty term to the loss function. It is used to address the problem of **multicollinearity** (high correlation among predictors) and to prevent **overfitting** in models with a large number of features or small datasets.
- Formula:
https://github.com/luumsk/NSU_ML/blob/main/Lectures/Day4_RidgeRegression.pdf

Ridge Regression

The main idea behind **Ridge Regression** is to find a **New Line** that doesn't fit the **Training Data** as well...



In other words, by starting with a slightly worse fit, **Ridge Regression** can provide better long term predictions.



When to use Ridge Regression

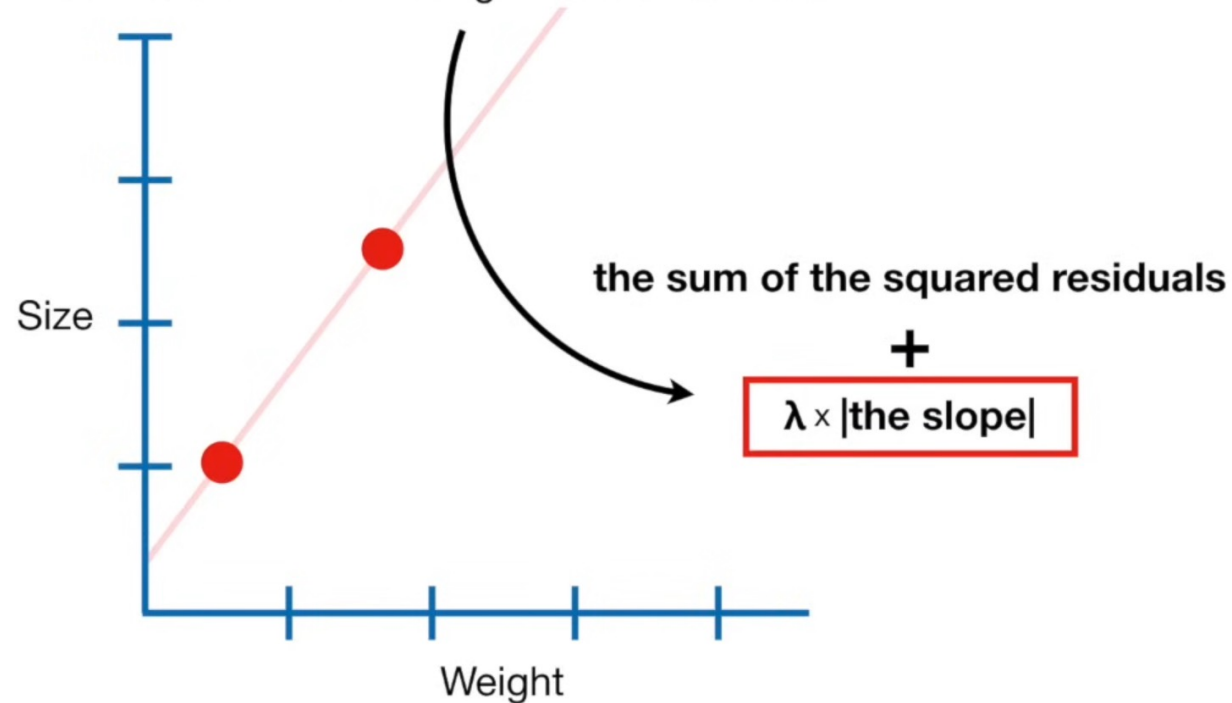
- You have **many correlated predictors**.
- You want to improve **generalization** without removing features.
- You are dealing with **small datasets** or **high-dimensional data**.

Lasso Regression

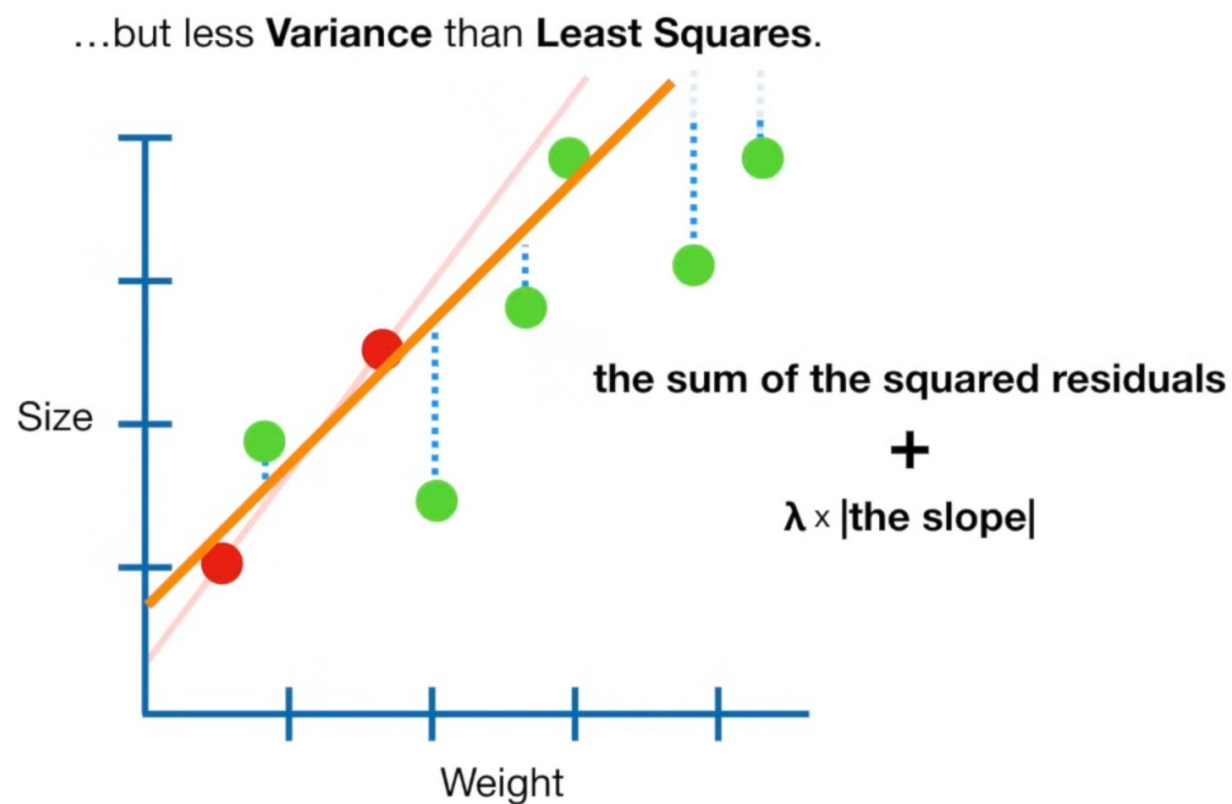
Lasso Regression

Least Absolute Shrinkage and Selection Operator

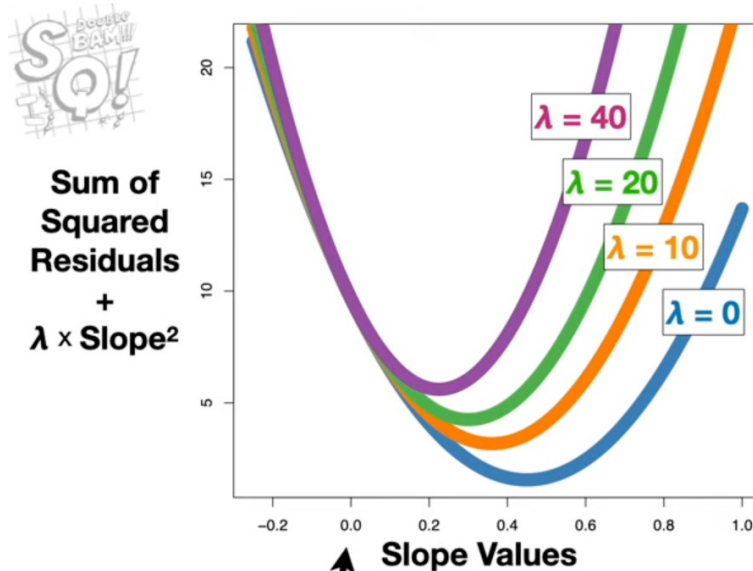
NOTE: Just like with **Ridge Regression**, λ can be any value from **0** to **positive infinity** and is determined using **Cross Validation**.



Lasso Regression

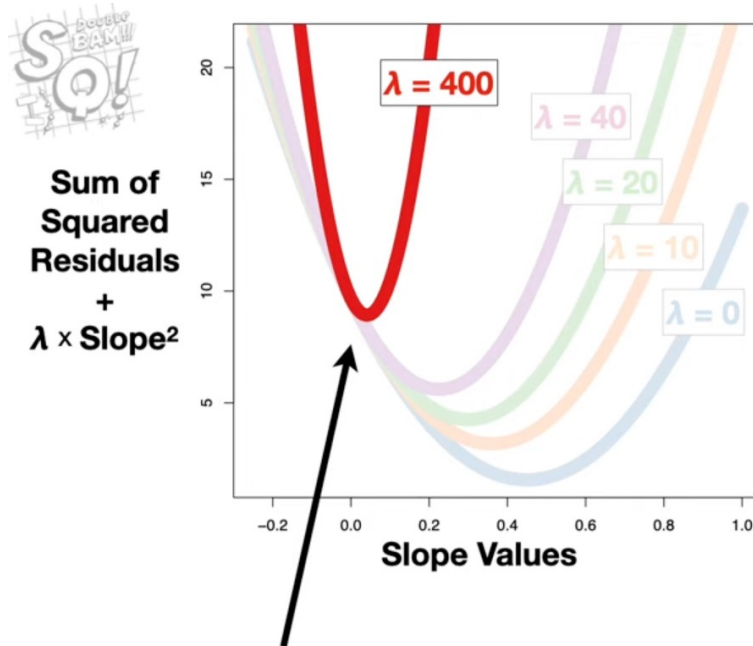


Ridge vs Lasso Regression



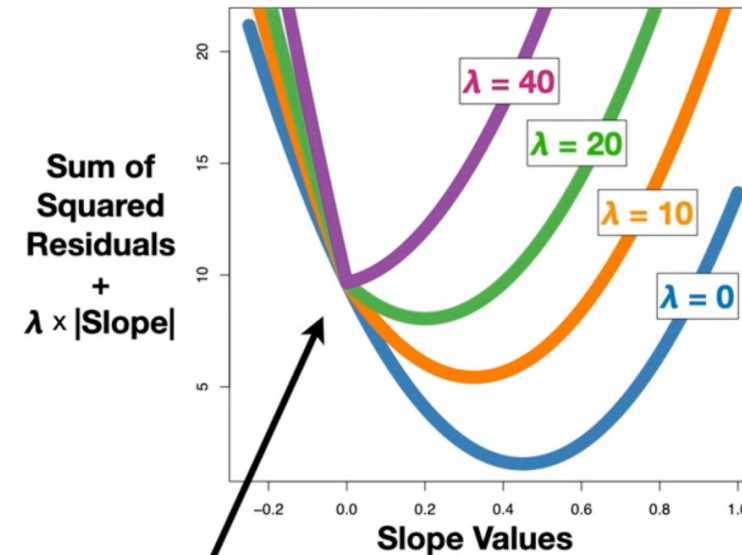
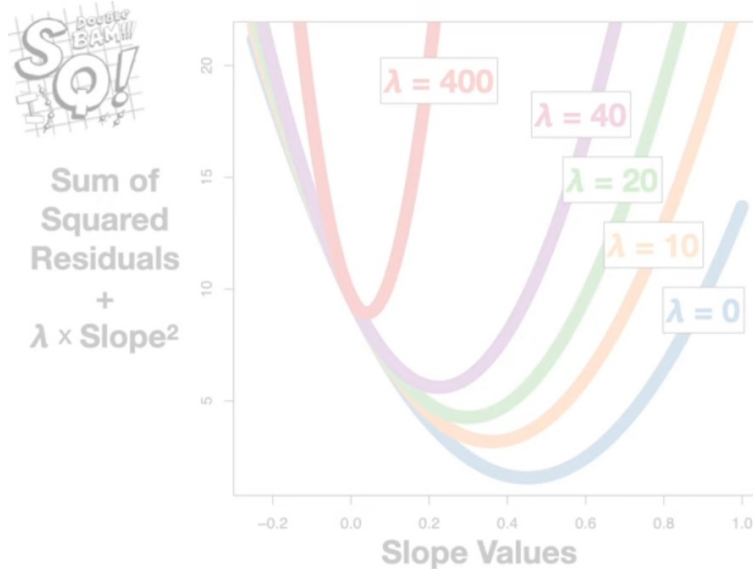
When we increase the
Ridge Regression Penalty,
aka the **L2 Penalty**, aka the
Square Penalty...

Ridge vs Lasso Regression



And even when we set λ (**lambda**) to something crazy high, like **400**, we still end up with an optimal slope > 0 .

Ridge vs Lasso Regression



...the optimal value shifts towards 0,
but, since we have a kink at 0, 0 ends
up being the optimal slope.

Data Visualization

Data Visualization

- Lab: <https://www.kaggle.com/code/khueluu/data-visualization>

Graded Assignment

- TBA
- Deadline: **Monday, 02.12.2024**

References

- <https://online.stat.psu.edu/stat462/node/158/>
- <https://tahera-firdose.medium.com/understanding-polynomial-regression-603eb25501d>
- <https://tahera-firdose.medium.com/understanding-polynomial-regression-603eb25501d>
- <https://www.cuemath.com/calculus/nonlinear-functions/>
- <https://www.coursera.org/learn/machine-learning-with-python?specialization=ibm-data-science>