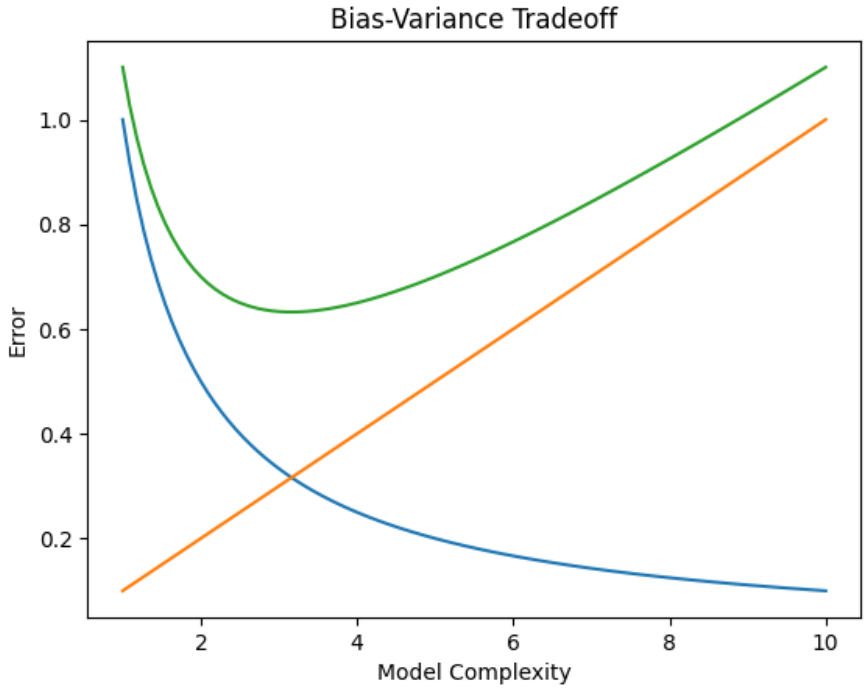
Machine Learning Concepts

1. What is the Bias–Variance Tradeoff?

It describes the balance between two types of errors:

- **Bias** → Error from overly simple models that miss patterns (underfitting)
- **Variance** → Error from models that are too sensitive to training data (overfitting)

A good model finds the right balance so it generalizes well to new data.



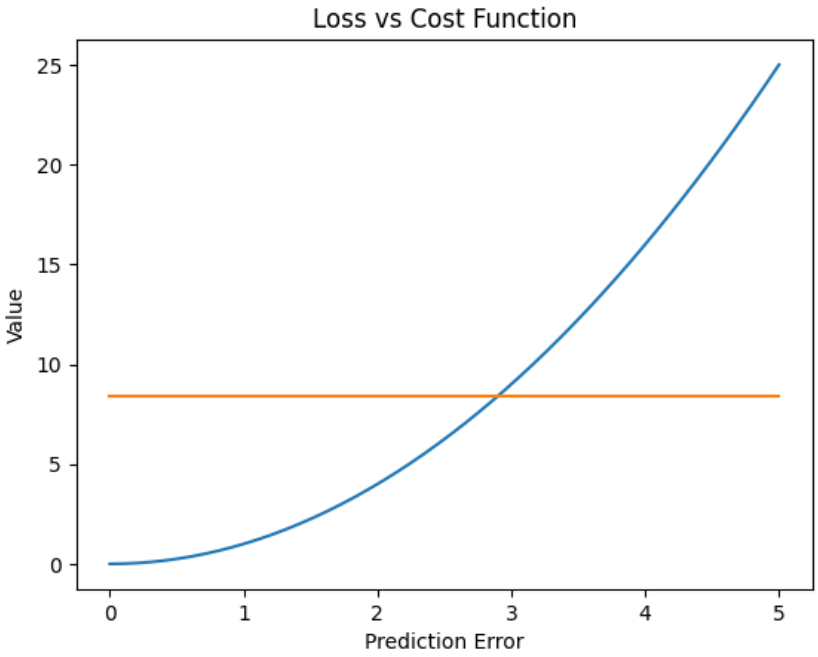
You saw three curves:

- **Bias** decreases as model complexity increases
- **Variance** increases with complexity
- **Total Error** is U-shaped → the sweet spot is in the middle

2. What is the difference between a loss function and a cost function?

- **Loss Function** → Measures error for one data point
- **Cost Function** → Average (or total) loss over the entire dataset

Loss = individual error, Cost = overall error.



- The curved line shows loss for individual errors
- The straight horizontal line shows overall cost (average loss)

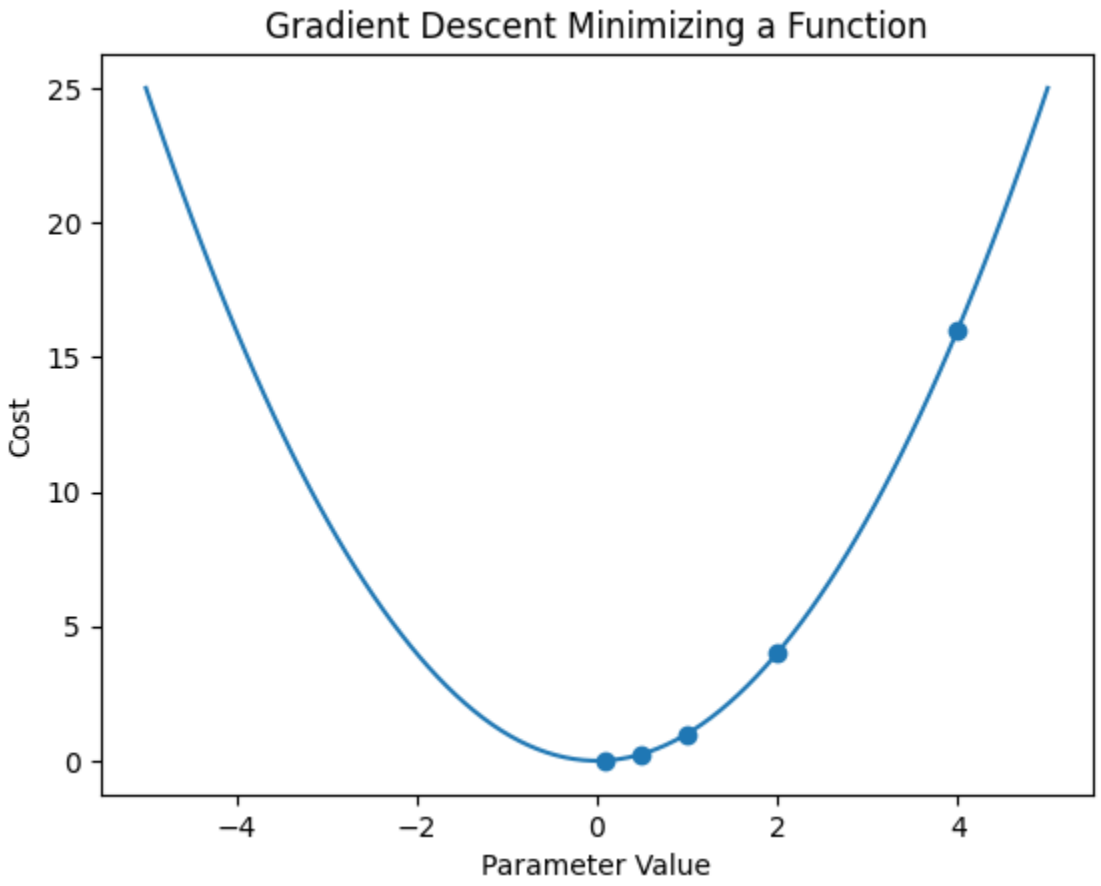
3. What is Gradient Descent?

Gradient Descent is an optimization algorithm used to minimize the cost function.

It works by:

1. Calculating the slope (gradient) of the cost function
2. Moving parameters in the opposite direction of the gradient
3. Repeating until the error is as small as possible

It's how many ML models learn the best parameter values.



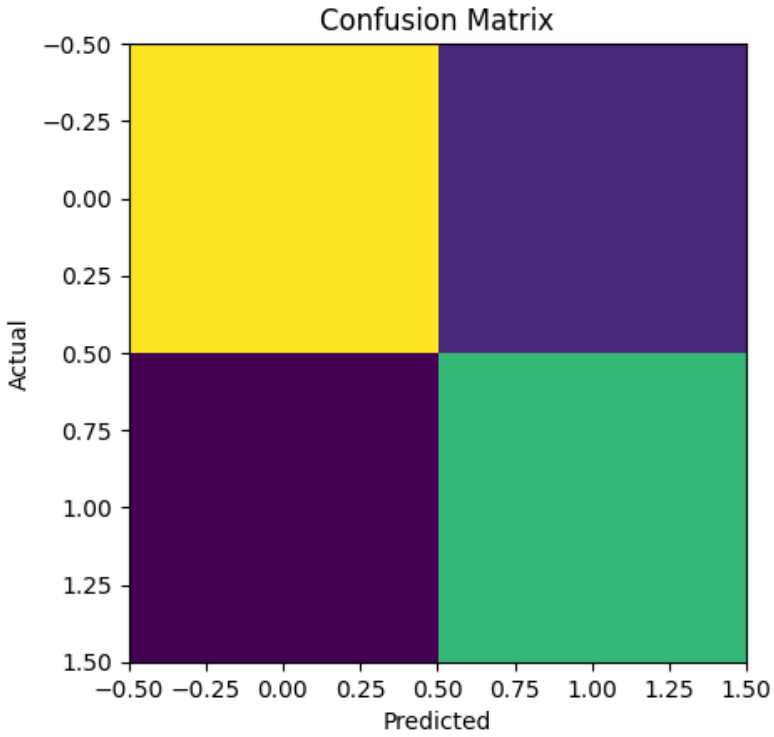
- The curve is the **cost function**
- The dots show steps moving **downhill toward the minimum**

4. What is a Confusion Matrix?

A confusion matrix is a table used to evaluate classification models.

	Predicted Positive	**Predicted Negative**
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

It helps calculate metrics like accuracy, precision, recall, and F1-score.



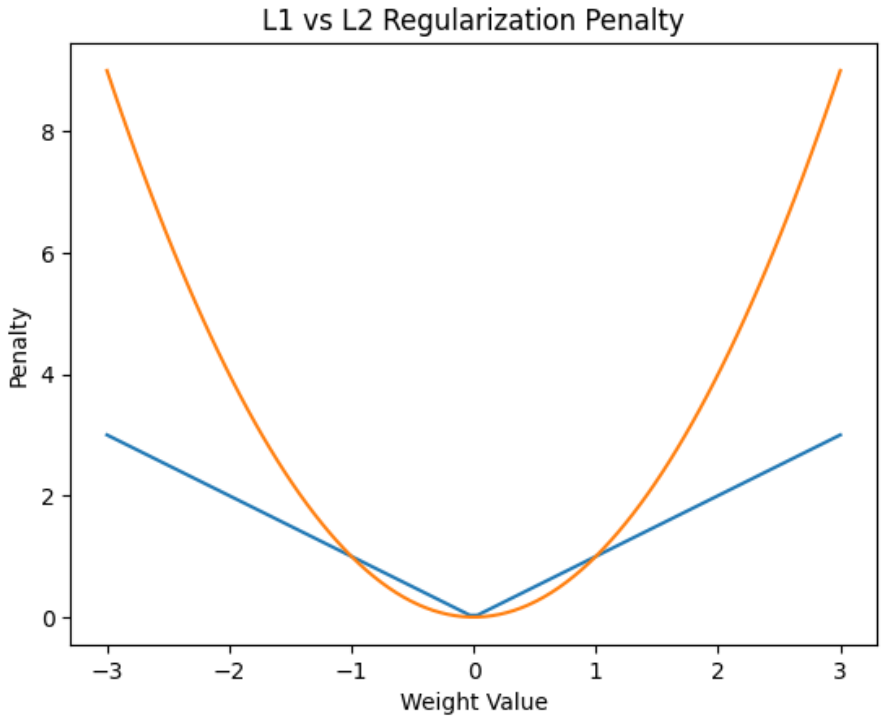
A 2×2 grid showing:

- Top-left → True Positives
- Top-right → False Positives
- Bottom-left → False Negatives
- Bottom-right → True Negatives

5. What are L1 (Lasso) and L2 (Ridge) Regularization?

Both techniques reduce overfitting by adding a penalty to large coefficients.

- **L1 (Lasso)** → Adds absolute values of coefficients
    - Can shrink some coefficients to zero → performs feature selection
- **L2 (Ridge)** → Adds squared values of coefficients
    - Shrinks coefficients but rarely makes them zero



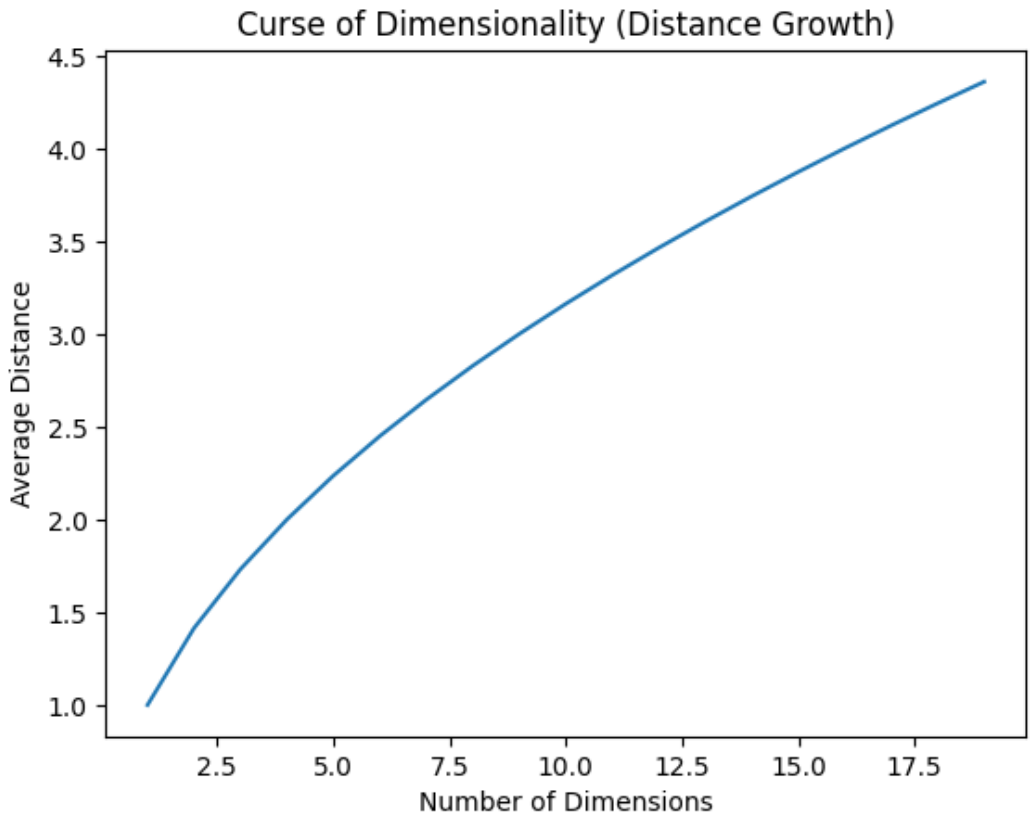
- **L1** forms a sharp V-shape → pushes some weights to zero
- **L2** forms a smooth curve → shrink weights but rarely zero

6. What is the “Curse of Dimensionality”?

As the number of features (dimensions) increases:

- Data becomes sparse
- Distance-based algorithms become less meaningful
- Models require much more data to learn well

High dimensions can make learning harder and less accurate.

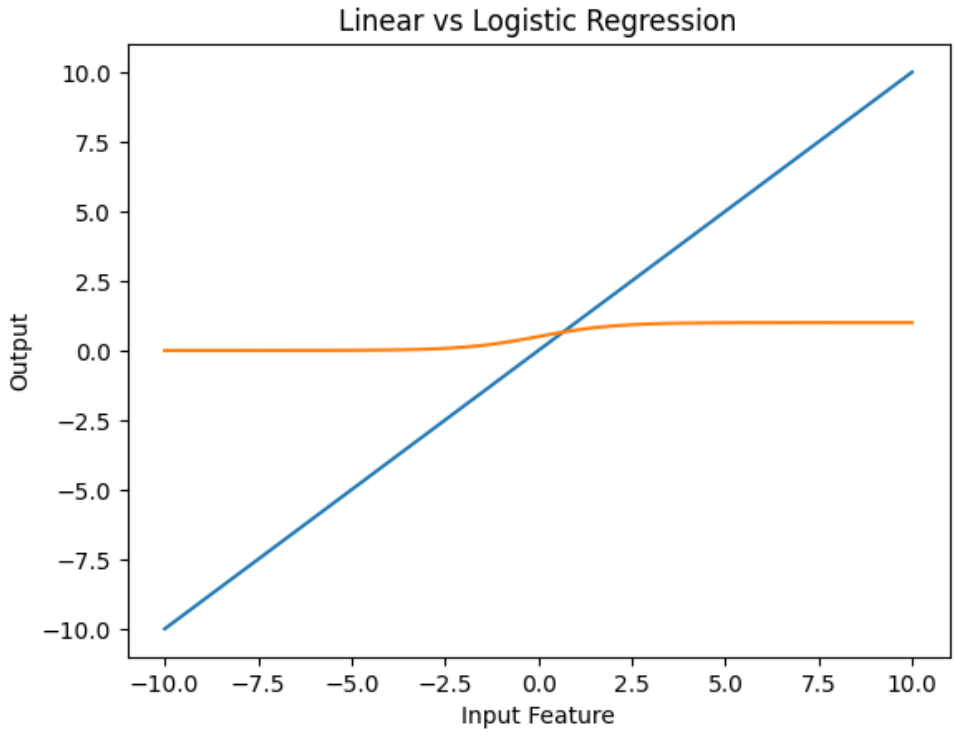


- As dimensions increase, average distance between points grows
- This makes learning harder and data more sparse

7. How does Logistic Regression differ from Linear Regression?

Linear Regression	**Logistic Regression**
Predicts continuous values	Predicts probabilities of classes
Output can be any number	Output is between 0 and 1
Uses straight line	Uses sigmoid function to map values to probabilities

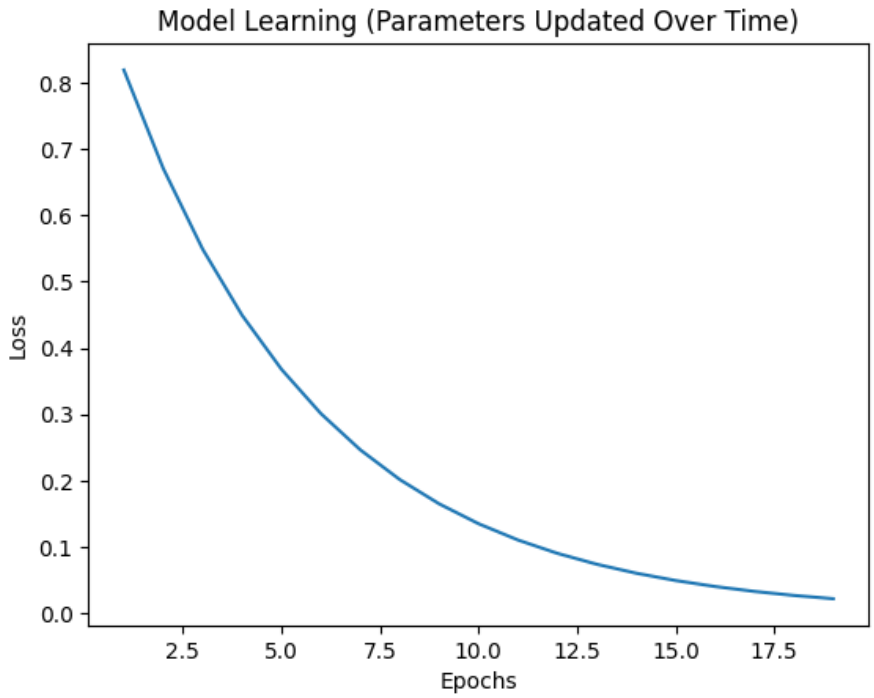
Logistic Regression is mainly used for classification.



- Straight line → Linear Regression output
- S-shaped curve → Logistic Regression probability output (0 to 1)

8. What is the difference between Parameters and Hyperparameters?

- **Parameters** → Learned automatically during training
  Example: weights in linear regression or neural networks
- **Hyperparameters** → Set before training by the user
  Example: learning rate, number of trees, number of layers

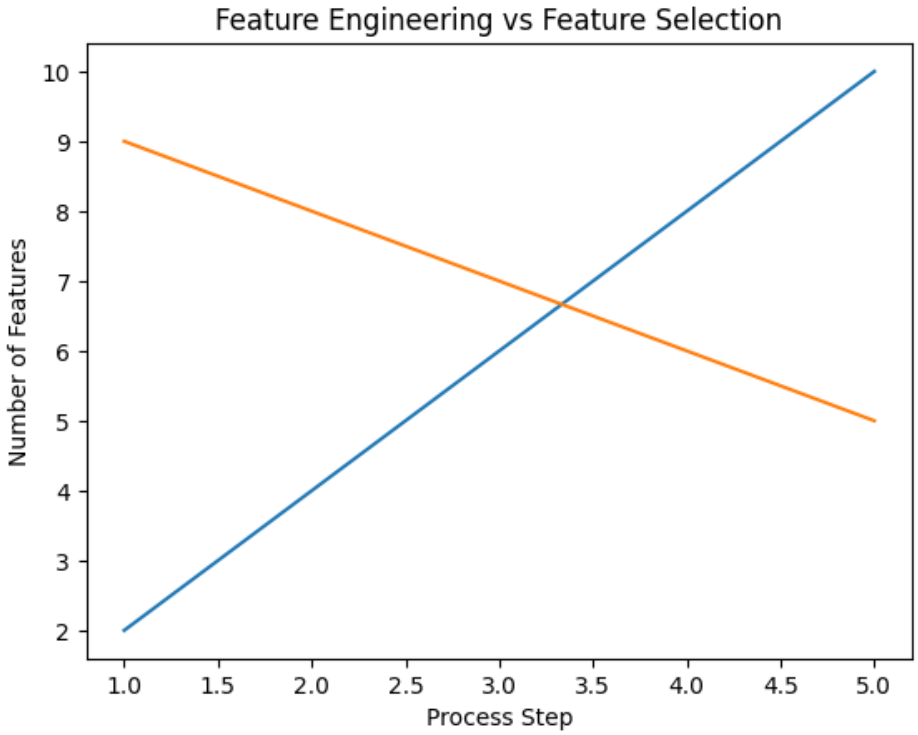


- Loss decreases as epochs increase
- Shows how parameters improve during training

9. What is the difference between Feature Engineering and Feature Selection?

- **Feature Engineering** → Creating new features from existing data
  Example: extracting day/month from a date
- **Feature Selection** → Choosing the most important features and removing irrelevant ones

Engineering adds, selection reduces.

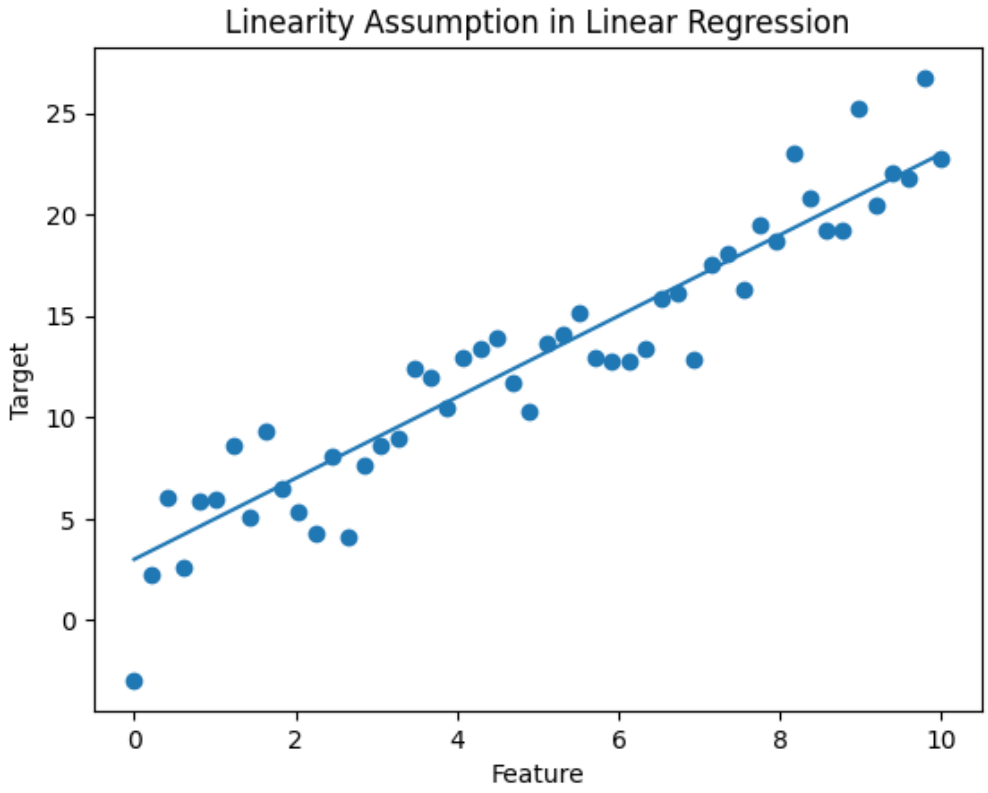


- One line increases → adding new features
- One line decreases → removing less useful features

10. What are the key assumptions of Linear Regression?

1. **Linearity** → Relationship between features and target is linear
2. **Independence** → Observations are independent
3. **Homoscedasticity** → Constant variance of errors
4. **Normality of Errors** → Residuals are normally distributed
5. **No Multicollinearity** → Features are not highly correlated

If these assumptions are violated, model performance can suffer.



- Scatter points follow a rough straight-line pattern
- The fitted line shows the assumed linear relationship