Underfitting:

Imagine you're trying to predict house prices with only the number of bedrooms, but you're ignoring location, area, etc.
Your model becomes too basic → **can't learn real patterns**.

Underfitting happens when your machine learning model is **too simple** to learn the underlying pattern of the data.

High Bias , Low Variance

- High error on **training** and **test** data
- Poor accuracy everywhere
- Model **under-learns**

Overfitting:

You train a model to identify cats and dogs, and it even learns the shadow, camera angle, or background color — things that don't actually help with prediction.
On test images, it fails miserably.

Overfitting happens when your model is **too complex** and learns **both the patterns and the noise** in the training data.

Low Bias , High Variance

- Excellent performance on **training data**
- Poor performance on **new/unseen (test) data**
- Model **memorizes instead of learning**

We want a model that fits **between underfitting and overfitting** — this is called a **well-generalized model**.

Generalized Model:

A **generalized model** is a machine learning model that:

- **Learns the true patterns** from training data,
- **Avoids memorizing noise** or random details,
- And performs **well on new, unseen data** (test data).

"A generalized model is smart enough to learn the real logic from the data — not too simple, not too complex — just right."

You train a model to predict house prices based on:

- Area
- Location
- Number of bedrooms

If your model learns the **real relationship** between these factors and price (without overfitting on exact training prices), then it's **generalized**.

✅ It will predict new house prices accurately, even for houses it has never seen before.

How to Build a Generalized Model?

1. ✅ Use **Cross-validation**
2. ✅ Apply **Regularization** (like Ridge/Lasso)
3. ✅ Use **sufficient training data**
4. ✅ Avoid irrelevant features (Feature selection)
5. ✅ Fine-tune **model complexity** (not too deep, not too shallow)

Bias: Bias means your model is **too simple** and is **not learning properly** from the data.

Imagine you're trying to guess the weight of a person just by looking at their **height only** (ignoring age, gender, or body type).
You will keep making wrong predictions because you're using **very limited knowledge** → this is **high bias**.

- Model assumes **wrong or simple rules**.
- Doesn't learn the full picture.
- **Underfitting** happens due to high bias.

Variance: Variance means your model is **too smart or too sensitive**, it learns even the **random noise or mistakes** in the training data.

Imagine you memorize all the questions in a book to prepare for an exam — but in the actual test, the questions are a little different.
You fail because you **memorized**, you didn't **understand the logic** → this is **high variance**.

- Model **reacts too much** to small changes in training data.
- It performs great on training data but poorly on new data.
- **Overfitting** happens due to high variance.

Ideal model = Low Bias + Low Variance

Ridge and Lasso Regression:

What is Ridge Regression?

1. Ridge Regression is used when **normal linear regression gives poor results**.
2. It works by **making the model simpler**.
3. It does this by **reducing the size of the coefficients** (the numbers in front of features).
4. It **does not remove** any features, just **reduces their impact**.
5. It is helpful when **all features (columns) are important**.
6. It helps to avoid **overfitting** (when model works well on training but badly on test data).
7. Also known as L2 Regularization.

What is Lasso Regression?

1. Lasso is also used to improve linear regression results.
2. It makes the model simple **by reducing and removing unnecessary features**.
3. It can **set some feature values to zero**, meaning it ignores them.
4. It helps in **feature selection** — choosing only the useful features.
5. It is helpful when **only some features are useful** and others are not.
6. It also helps avoid **overfitting**.
7. Also known as L1 Regularization.

Imagine you are predicting house price using:

- Area
- Bedrooms
- Color of the door
- Direction the window faces

Now:

- **Ridge** will keep **all** features but reduce the effect of less useful ones.
- **Lasso** will completely **remove** features like door color and window direction if they are not helping.

What is R^2

Score? R^2 (R-squared) is a **performance metric** used to check how well your **regression model** is working.

R^2 Score Value Range:

- **$R^2 = 1$:** Perfect fit (model predicts exactly)
- **$R^2 = 0$:** Model predicts nothing better than average
- **$R^2 < 0$:** Model is worse than a dummy model (bad model)

💡 Simple Example:

Suppose you're predicting **house price** using **area**.

- If $R^2 = 0.90$ → your model explains **90% of the variation** in house prices.
- If $R^2 = 0.20$ → your model explains only **20%**. That means it's a poor fit.
- If $R^2 = 1.0$ → perfect prediction (very rare in real life).