🌈1️⃣What is Softmax Regression?

Softmax regression (also called **multinomial logistic regression**) is a **classification model** used when you have **more than two possible classes**.

💡 Logistic regression → 2 classes (binary: spam / not spam).
💡 Softmax regression → 3 or more classes (multiclass: spam / personal / work mail).

So, softmax regression is just a **generalization** of logistic regression.

🎯2️⃣What does the model do?

The goal of softmax regression is to predict **the probability** that an input xxx belongs to each of the kkk classes.

It gives you probabilities like:

- P(y=1∣x)=0.7P(y = 1 | x) = 0.7P(y=1∣x)=0.7
- P(y=2∣x)=0.2P(y = 2 | x) = 0.2P(y=2∣x)=0.2
- P(y=3∣x)=0.1P(y = 3 | x) = 0.1P(y=3∣x)=0.1

And since these are probabilities:

P(y=1∣x)+P(y=2∣x)+P(y=3∣x)=1P(y=1|x) + P(y=2|x) + P(y=3|x) = 1P(y=1∣x)+P(y=2∣x)+P(y=3∣x)=1

⚙️3️⃣How does it work?

We assume that each class iii has its own **parameter vector** θi\theta_iθi.

For an input xxx:

- Compute a **score** for each class:

$$s_i = \theta_i^T x$$

- Then, apply the **softmax function** to convert those scores into probabilities:

$$P(y = i|x) = \frac{e^{s_i}}{\sum_{j=1}^{k} e^{s_j}}$$

This formula ensures:

- All probabilities are **positive**.
- They **sum to 1**.

🧩4️⃣ The link with logistic regression

If k=2k = 2k=2, the softmax formula becomes the same as **logistic regression**.

That's why we say:

> "Softmax regression generalizes logistic regression."

📘5️⃣ How do we find the parameters θi\theta_iθi?

We **train** the model using **maximum likelihood estimation** —
which means we find the values of θi\theta_iθi that **maximize the probability of the observed data**.

To do this, we:

1. Write the **log-likelihood** function (how likely our data is under the model).
2. Use **gradient ascent** (or similar methods) to find the best parameters.

You don't need to derive this by hand — in practice, libraries like **Scikit-learn** or **TensorFlow** do it for you.