**HW to Chapter 12 “Softmax”**

**Non-programming Assignment**

1. What is the reason for softmax?

Softmax is used in multiclass classification problems where the goal is to estimate the probability of each class. Instead of outputting a binary result, softmax converts raw model outputs (logits) into a normalized probability distribution over K possible classes. This allows us to interpret the model's output as the probability of a data point belonging to each class, with all probabilities summing up to 1.

1. What is softmax and how does it works?

Softmax is an activation function used in the final layer of neural networks when performing multiclass classification. It takes a vector of values (logits) and transforms them into probabilities by applying the following formula:

softmax(zj​)= e(zj) ​​/ ∑K k=1​ e(zk)

Where zj the raw output (logit) for class j, and K is the number of possible classes. This function ensures that the output is a probability distribution: each value is between 0 and 1, and the sum of all values is 1.