

# Classification with Logistic Regression

Explain sigmoid function?

In this breast cancer classification task, we used logistic regression to predict whether a tumor is benign (0) or malignant (1).

Logistic regression first calculates a linear combination of the input features:

$$Z = W_0 + W_1X_1 + W_2X_2 + \dots$$

The sigmoid function then transforms this z value into a probability between 0 and 1:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

- If the probability is close to 1, the model is confident that the tumor is malignant.
- If the probability is close to 0, the model is confident that the tumor is benign.

By default, the decision threshold is 0.5:

- Probability  $\geq 0.5 \rightarrow$  predict malignant (1)
- Probability  $< 0.5 \rightarrow$  predict benign (0)

In order to understand how predictions change, the threshold was tuned to 0.3.

Lowering the threshold makes the model classify more tumors as malignant, which increases recall (catching more true malignant cases) but may reduce precision (more false positives).

The sigmoid function is what makes this possible. Without it, logistic regression couldn't output probabilities for flexible threshold adjustment.