INMAS

# Logistic Regression

James Balamuta

INMAS Statistical Methods Workshop Fall 2021

# Lecture Objectives

- *Relate* linear regression and logistic regression
- *Emphasis* difference between estimating probabilities and quantities.

# Logistic Regression

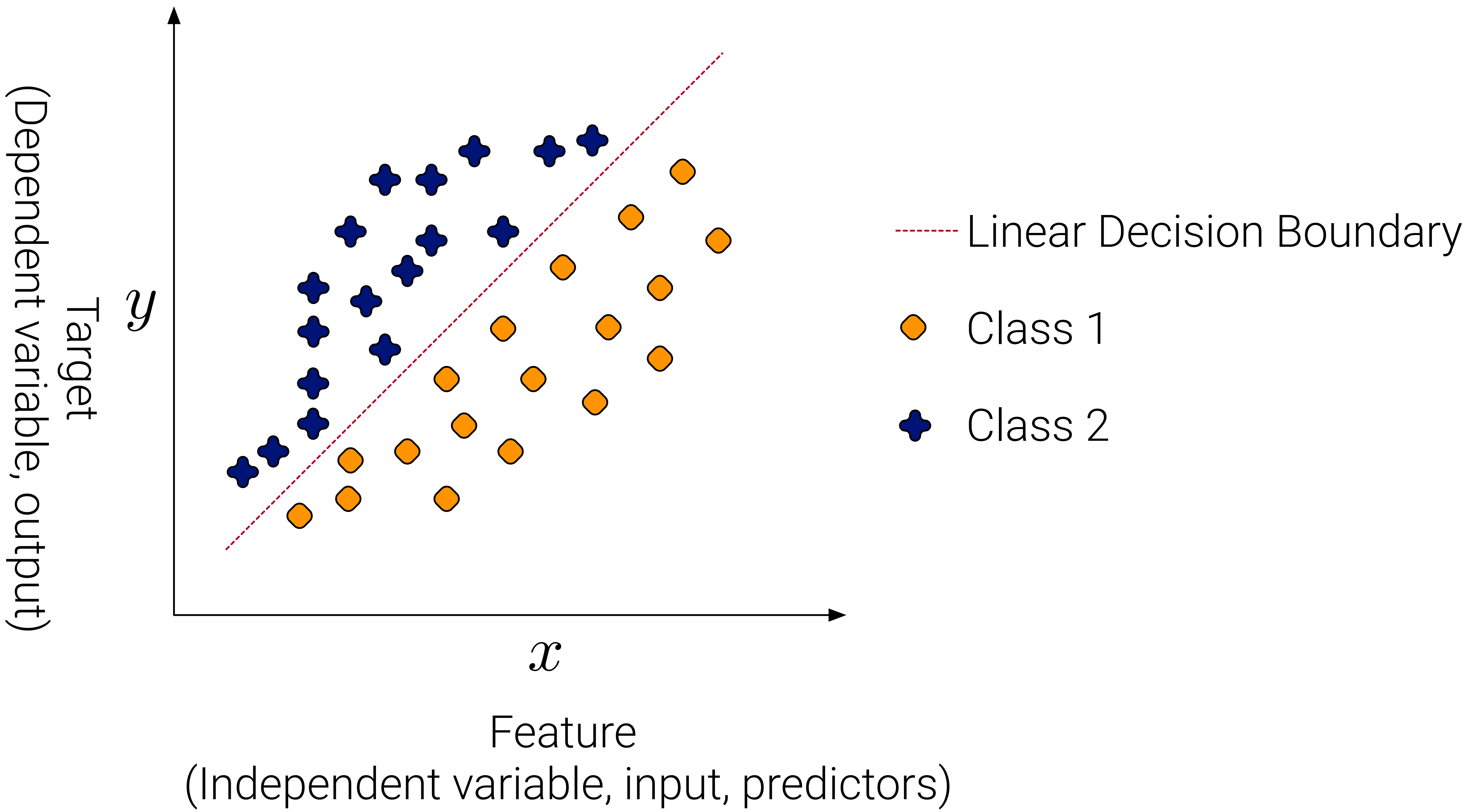
Previously

# Learning Types

- **Supervised Learning** with Labeled Data. (Today)
  - Methods: Regression or classification
  - Objective: To predict a response or **outcome**.
- **Unsupervised Learning** with Unlabeled Data.
  - Methods: Clustering, Principal Component Analysis (PCA), autoencoders, generative adversarial networks (GANs)
  - Objective: Identify patterns in the data or understand how data was created.
- Best distinction between the two:
  - **Is there a response variable Y?**

Previously

# **Supervised Learning**: Classification



# How can we classify the data?

Aside

# Binary (Logical) Data

Data exists in two states: present or absent

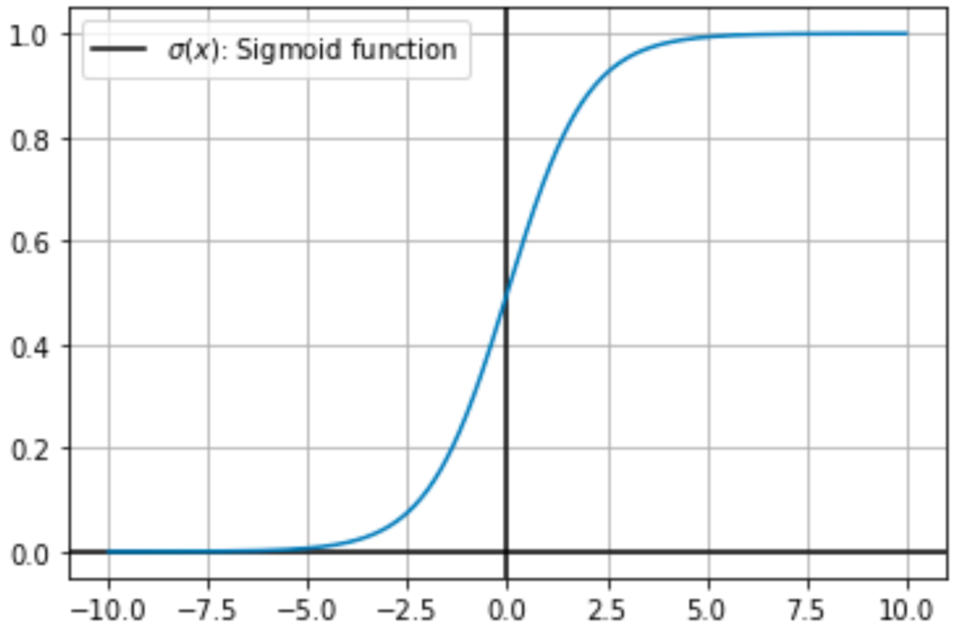
| | | |
|---|---|---|
| TRUE | T | (1) |
| FALSE | F | (0) |

Source

* Logical values allow us to determine whether a condition is met or not.
Based on the condition, we can make a **choice** as to what happens next.

# Sigmoid

## Scope of Probability

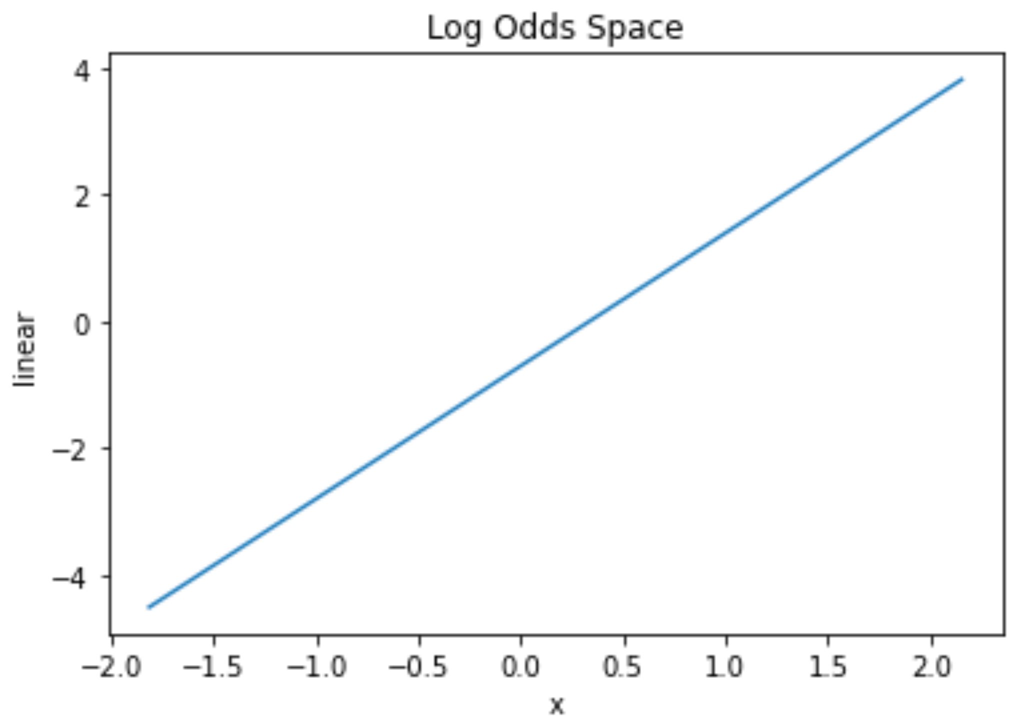


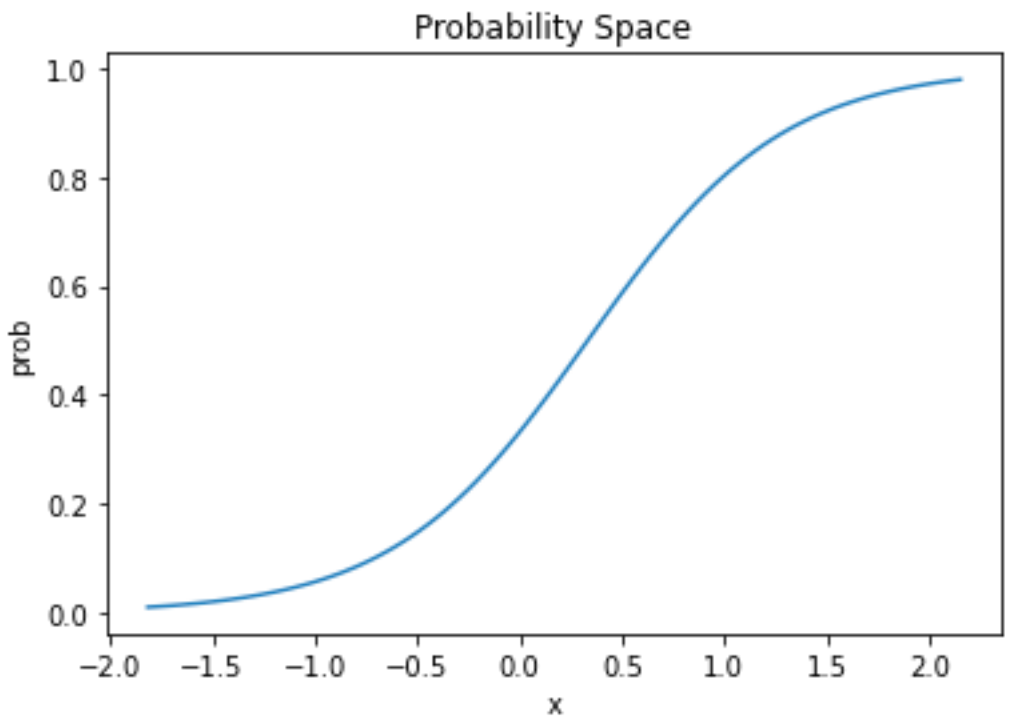

$$g(z) = \sigma(z) = \frac{\exp(z)}{1+\exp(z)} = \frac{1}{1+\exp(-z)}$$

What happens when ***z*** is:

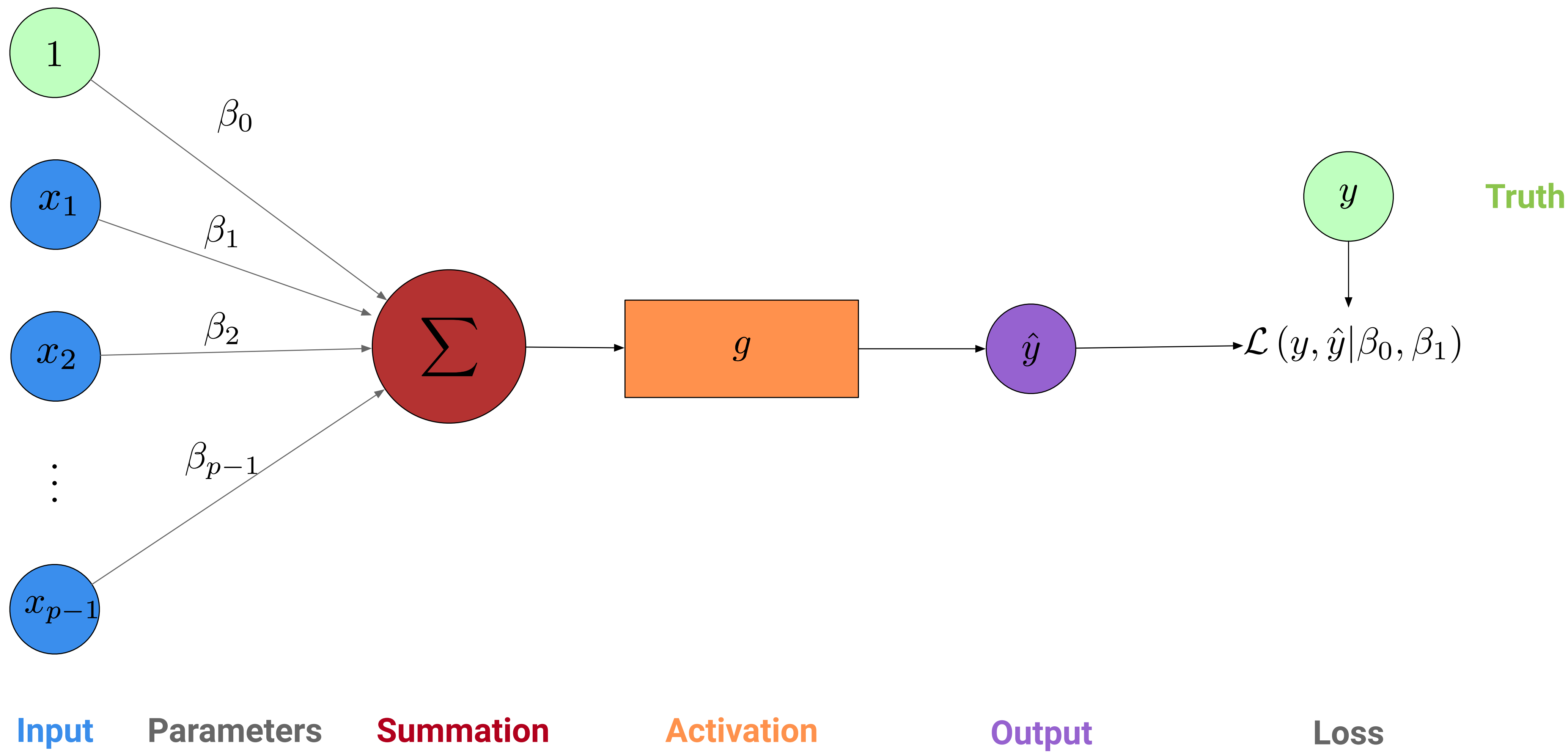
- a large positive number?
- a large negative number?

# Going Between Spaces



# Logistic Regression



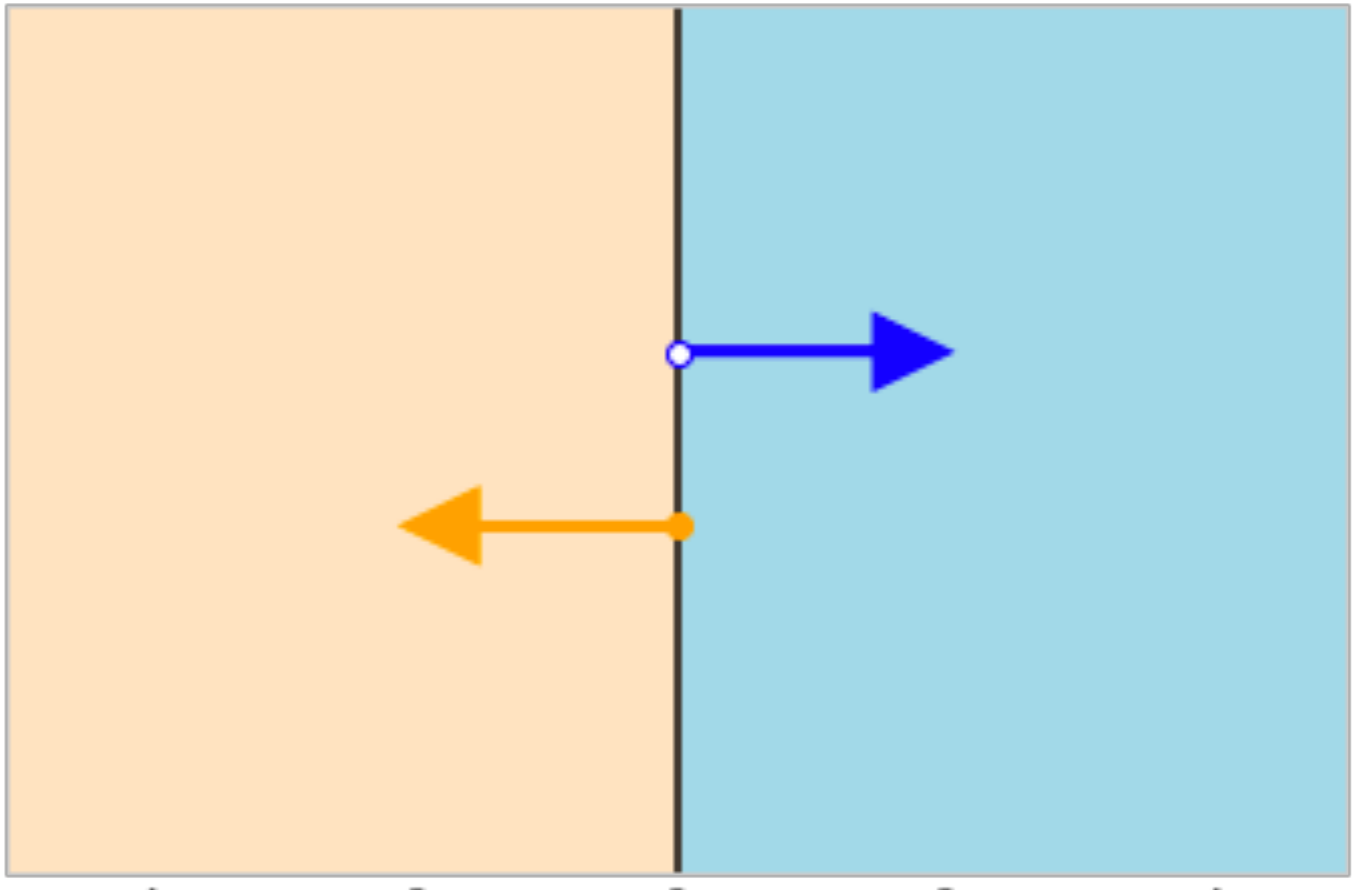

$x_i \in \mathbb{R}$

$$z = \sum_{j=0}^{p-1} \beta_j x_j \qquad g(z) = \frac{1}{1 + \exp\left(-z\right)} \qquad 0 \leq \hat{y} \leq 1 \qquad \mathcal{L}(y, \hat{y}) = -\left[y \log\left(\hat{y}\right) + (1-y) \log\left(1-\hat{y}\right)\right]$$

Binary Cross-Entropy

# Threshold

Applying a cut-off value to make a decision



# Logistic Regression
## Algorithmic View

1. Initialize parameters to zero, e.g. $\boldsymbol{\beta} := 0$
2. Under each training epoch:

Compute for each sample $\langle \boldsymbol{x}^{(i)}, y^{(i)} \rangle \in \mathcal{D}$

A prediction: $\hat{y}^{(i)} := \sigma(\boldsymbol{x}^{(i),\top}\boldsymbol{\beta})$

Prediction error: $\mathrm{error} := y^{(i)} - \hat{y}^{(i)}$

Gradients $\nabla_{\boldsymbol{\beta}}\mathcal{L} := -(y^{(i)} - \hat{y}^{(i)})\boldsymbol{x}^{(i)}$

Parameter update: $\boldsymbol{\beta} := \boldsymbol{\beta} + \eta \cdot \nabla_{\boldsymbol{\beta}}\mathcal{L}$

# Summary

- Logistic Regression was taken to be better approach for classification.
- Logistic regression still produces a "quantity" but we apply a threshold decision to it.