CMPINF 2100

Introduction to Data Centric Computing

Week 11

Logistic – what is the logit??

Binary classification

- Binary classifiers CLASSIFY the **EVENT** or the **NON-EVENT**.

- The **EVENT** is commonly referred to as `y=1`.

- The **NON-EVENT** is commonly referred to as `y=0`.

Binary classification

- Binary classifiers CLASSIFY the **EVENT** or the **NON-EVENT**.
- The **EVENT** is commonly referred to as `y=1`.
- The **NON-EVENT** is commonly referred to as `y=0`.
- The Binary OUTPUT, `y`, is therefore an INTEGER data type!!!

Binary classification

- HOWEVER, this is NOT a regression problem!!!
- The OUTPUT is a NUMBER but there are ONLY 2 unique values!!!!
- Regression is appropriate when the OUTPUT has many ALLOWABLE numeric values!!!

Binary classification – Cannot work with LINEAR MODELS!!!!!

- Remember the ASSUMPTIONS of the LINEAR MODEL!!!!

- The OUTPUT is Normally distributed around the AVERAGE OUTPUT (trend)!!!

Binary classification – Cannot work with LINEAR MODELS!!!!!

- Consider the case with a SINGLE input linearly related to the AVERAGE OUTPUT:

$$y_n \mid \mu_n, \sigma \sim \mathrm{normal}(y_n \mid \mu_n, \sigma)$$

$$\mu_n = \beta_0 + \beta_1 \times x_n$$

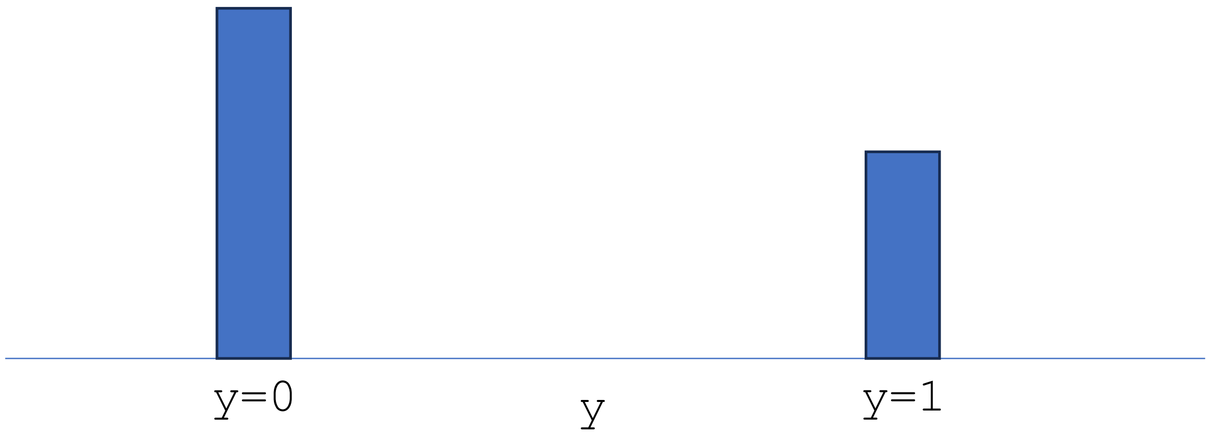
Binary classification – Cannot work with LINEAR MODELS!!!!!

- Consider the case with a SINGLE input linearly related to the AVERAGE OUTPUT:

$$y_n \mid \mu_n, \sigma \sim \text{normal}(y_n \mid \mu_n, \sigma)$$

The OUTPUT is BINARY.

It has 2 and ONLY 2 unique values!!!!



Where would the GAUSSIAN be located????

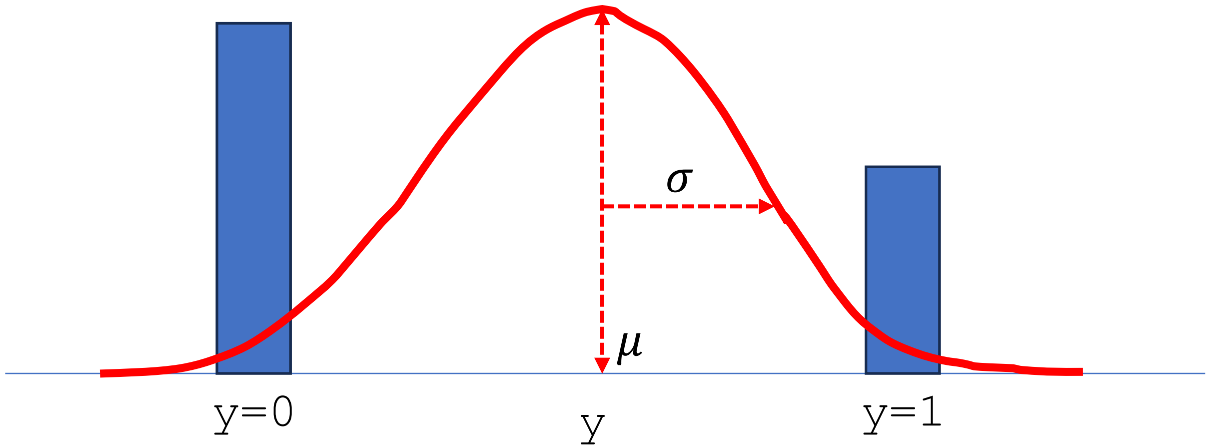
Binary classification – Cannot work with LINEAR MODELS!!!!!

- Consider the case with a SINGLE input linearly related to the AVERAGE OUTPUT:

$$y_n \mid \mu_n, \sigma \sim \text{normal}(y_n \mid \mu_n, \sigma)$$

The OUTPUT is BINARY.

It has 2 and ONLY 2 unique values!!!!



The OUTPUT is ONLY `y=0` OR `y=1`.
The AVERAGE cannot be in the MIDDLE!
The model CANNOT predict 0.5!!!

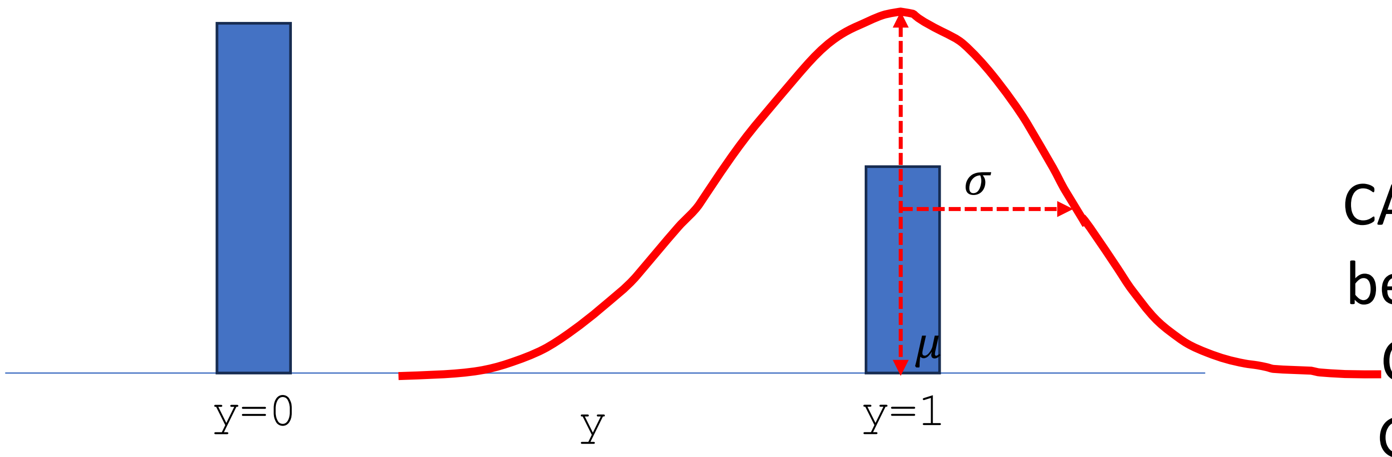
Binary classification – Cannot work with LINEAR MODELS!!!!!

- Consider the case with a SINGLE input linearly related to the AVERAGE OUTPUT:

$$y_n \mid \mu_n, \sigma \sim \text{normal}(y_n \mid \mu_n, \sigma)$$

The OUTPUT is BINARY.

It has 2 and ONLY 2 unique values!!!!



The OUTPUT is ONLY `y=0` OR `y=1`.
The AVERAGE CANNOT be at `y=1` because the model CANNOT predict GREATER than 1.

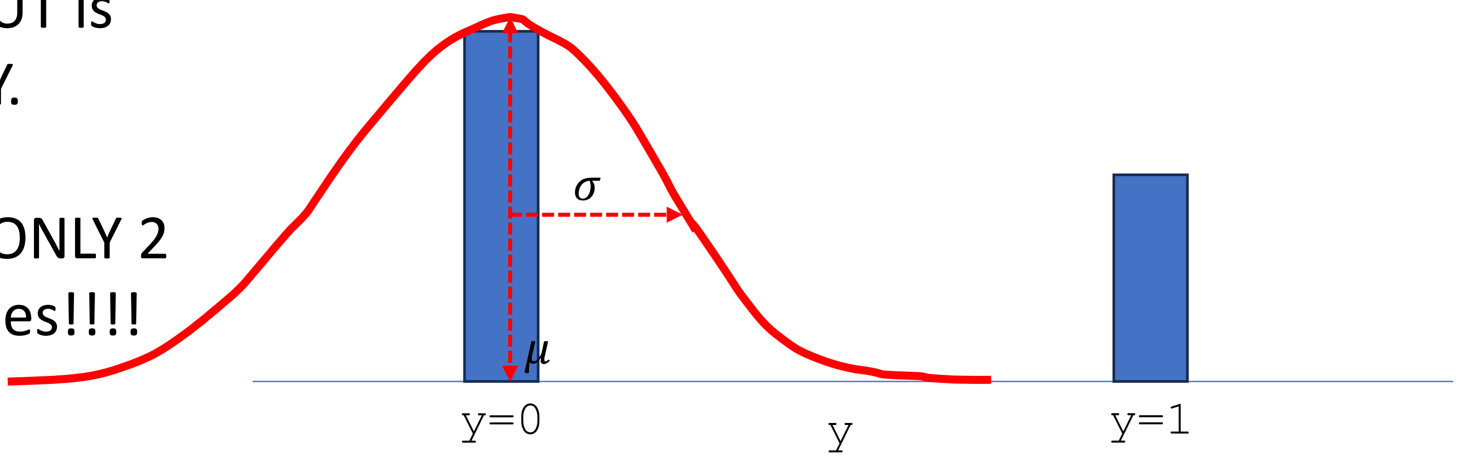
Binary classification – Cannot work with LINEAR MODELS!!!!!

- Consider the case with a SINGLE input linearly related to the AVERAGE OUTPUT:

$$y_n \mid \mu_n, \sigma \sim \text{normal}(y_n \mid \mu_n, \sigma)$$

The OUTPUT is BINARY.

It has 2 and ONLY 2 unique values!!!!



The OUTPUT is ONLY `y=0` OR `y=1`.
The AVERAGE CANNOT be at `y=0` because the model CANNOT predict LESS THAN 0.

Binary classification – Therefore does NOT use a Gaussian!!!

- The OUTPUT is NOT Normally distributed the AVERAGE OUTPUT!

- Instead, a different probability distribution is used!

- The OUTPUT is Bernoulli distributed around the AVERAGE OUTPUT!

$$y_n \mid \mu_n \sim \text{Bernoulli}(y_n \mid \mu_n)$$

Binary classification – Therefore does NOT use a Gaussian!!!

- The OUTPUT is NOT Normally distributed the AVERAGE OUTPUT!
- Instead, a different probability distribution is used!
- The OUTPUT is Bernoulli distributed around the AVERAGE OUTPUT!

$$y_n \mid \mu_n \sim \mathrm{Bernoulli}(y_n \mid \mu_n)$$

The AVERAGE OUTPUT or TREND has a special meaning!!!

The AVERAGE OUTPUT is the **<u>EVENT PROBABILITY</u>**!!!!!!

Binary classification models therefore PREDICT the EVENT PROBABILITY!!!!!

Why does this matter?

- Probabilities are BOUND between 0 and 1.
  - The EVENT PROBABILITY cannot be NEGATIVE.
  - The EVENT PROBABILITY cannot be GREATER THAN 1.

- This impacts how we model Binary Classification problems!!!!

Why does this matter?

- Return to the case of a single input LINEARLY related to the AVERAGE OUTPUT (trend):

$$\mu = \beta_0 + \beta_1 \times x$$

- The above equation allows for NEGATIVE TREND values!!!

- The above equation allows for TRENDS greater than 1!!!

- There’s NOTHING in the equation itself that KEEPS the TREND within the BOUNDS of a probability!!!!

Binary Classification – LOG ODDS RATIO

- Therefore, we CANNOT directly model the EVENT PROBABILITY!

- Instead, we must apply a TRANSORMATION to the AVERAGE OUTPUT.

- A popular approach is to MODEL or REGRESS the LOG ODDS RATIO.

$$\text{logodds} = \beta_0 + \beta_1 \times x$$

What is the LOG ODDS RATIO???

- The Log Odds Ratio is the NATURAL LOGARITHM of the ODDS RATIO (OR).

- The OR is defined as the PROBABILITY divided by 1 minus the PROBABILITY:

$$\mathrm{OR} = \frac{\mathrm{Probability}}{1 - \mathrm{Probability}}$$

What is the LOG ODDS RATIO???

- We defined the EVENT PROBABILITY as μ, thus the Odds-Ratio for the Binary classification problem is:

$$\mathrm{OR} = \frac{\mu}{1-\mu}$$

- The LOG ODDS RATIO is the NATURAL LOG of the OR:

$$\log(\mathrm{OR}) = \log\left(\frac{\mu}{1-\mu}\right)$$

The "best-fit-line" is applied to the LOG ODDS RATIO!!!!

- We model the Log-odds ratio instead of directly modeling the EVENT PROBABILITY!

$$\log(\mathrm{OR}) = \log\left(\frac{\mu}{1-\mu}\right) = \beta_0 + \beta_1 \times x$$

The "best-fit-line" is applied to the LOG ODDS RATIO!!!!

- We model the Log-odds ratio instead of directly modeling the EVENT PROBABILITY!

$$\boxed{\log(\mathrm{OR}) = \log\left(\frac{\mu}{1-\mu}\right)} = \beta_0 + \beta_1 \times x$$

The LOG-ODDS is also known as the **<u>LOGIT</u>**!!!!!!! → We are regressing the LOGIT of the EVENT PROBABILITY

The “best-fit-line” is applied to the LOG ODDS RATIO!!!!

- We model the Log-odds ratio instead of directly modeling the EVENT PROBABILITY!

$$\log(\mathrm{OR}) = \log\left(\frac{\mu}{1-\mu}\right) = \mathrm{logit}(\mu) = \beta_0 + \beta_1 \times x$$

We are regressing the **LOGIT** of the **EVENT PROBABILITY**

After fitting, the EVENT PROBABILITY is calculated by INVERTING the LOGIT

- The INVERSE LOGIT is known as the LOGISTIC!!!

- **Logistic regression** gets its name from the INVERSE of the LOGIT!!!!

$$\mu = \mathrm{INVERSE}\big(\mathrm{logit}(\beta_0 + \beta_1 \times x)\big) = \mathrm{logit}^{-1}(\beta_0 + \beta_1 \times x)$$

Important assumptions to remember with LOGISTIC REGRESSION

- Although REGRESSION is in the name, this is NOT a regression method for continuous outputs!
  - Logistic regression is for BINARY CLASSIFICATION!!!

- The model still predicts the AVERAGE OUTPUT, but the AVERAGE corresponds to the EVENT PROBABILITY!!!

- The LOGIT or LOG ODDS RATIO transformation is applied to make sure the EVENT PROBABILITY is bounded between 0 and 1!!!

- The transformation allows us to use essentially everything else from the LINEAR MODEL!
  - Logistic regression is a type of Generalized Linear Model (GLM)!!!!