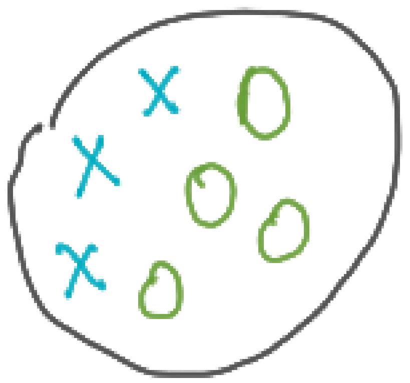


$$\text{Odds that } \underline{\text{I pick a X}} \text{ randomly} = \frac{\overset{\to\ 3/7}{P(\text{I pick X})}}{\underset{\to\ 4/7}{P(\text{I do NOT pick X})}} = \frac{3/7}{4/7} = \frac{3}{4}$$

$$P(\text{event}) = p \rightarrow [0,1]$$

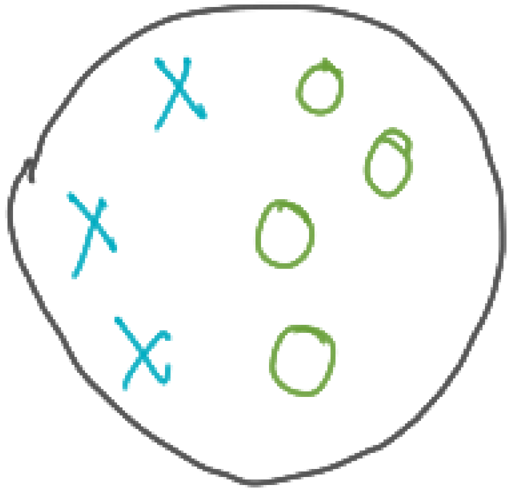
$$\boxed{\text{Odds}(\text{event}) = \frac{p}{1-p}} \rightarrow [0,\infty)$$

$$P(X) = \frac{3}{7} = p$$

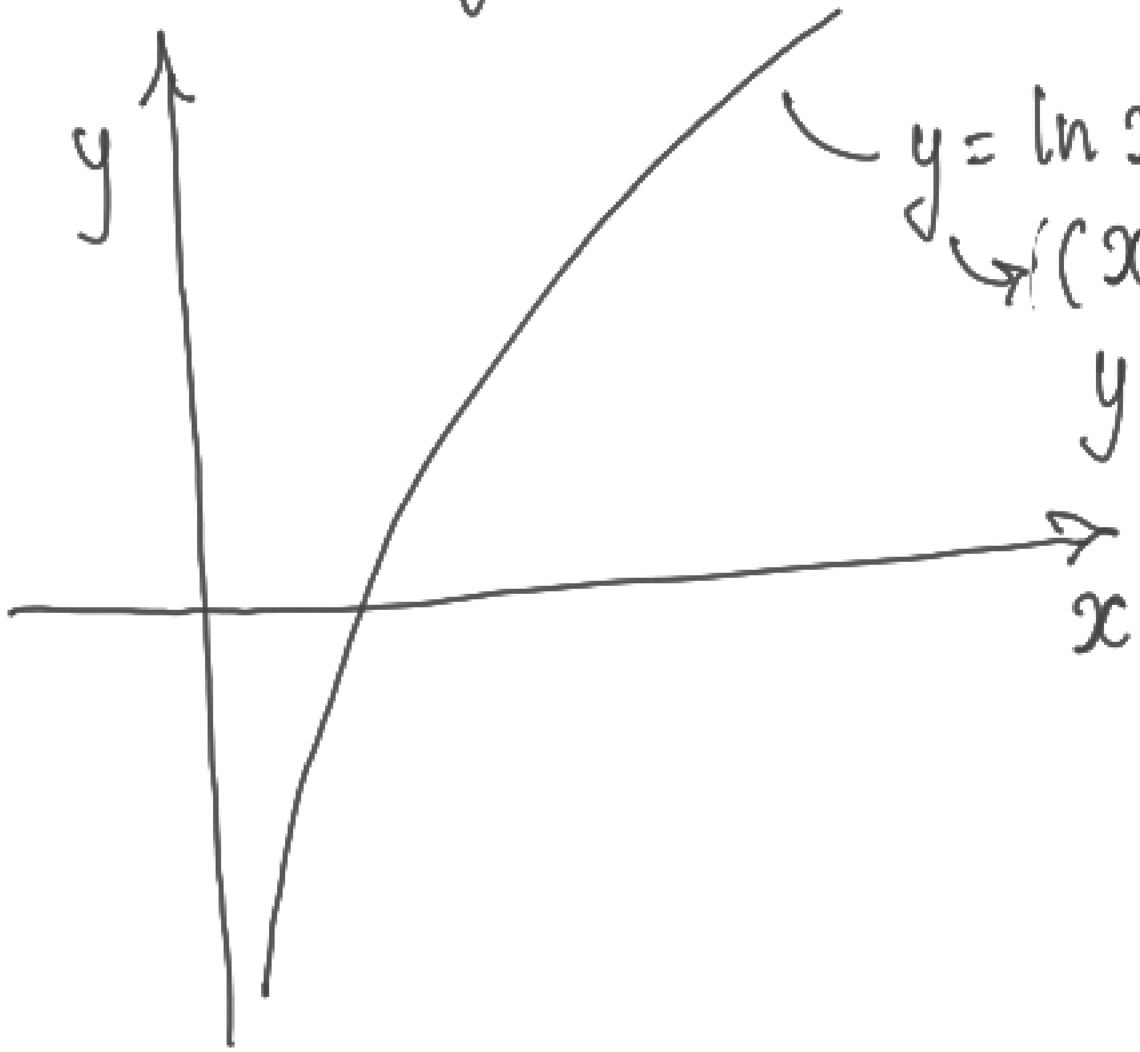
$$\text{Odds}(X) = \frac{p}{1-p} = \frac{3/7}{\underbrace{1-3/7}_{4/7}} = \frac{3}{4}$$

$$\text{Log Odds (event)} = \underbrace{\log}_{\ln}(\text{Odds (event)}) \rightarrow (-\infty, \infty)$$

$e \sim 2.74$



$$\text{Log Odds (I pick X randomly)} = \ln\left(\frac{3}{4}\right) = \log_e\left(\frac{3}{4}\right)$$



$y = \ln x$

$(x > 0)$

$y \in (-\infty, \infty)$

Why Log Odds

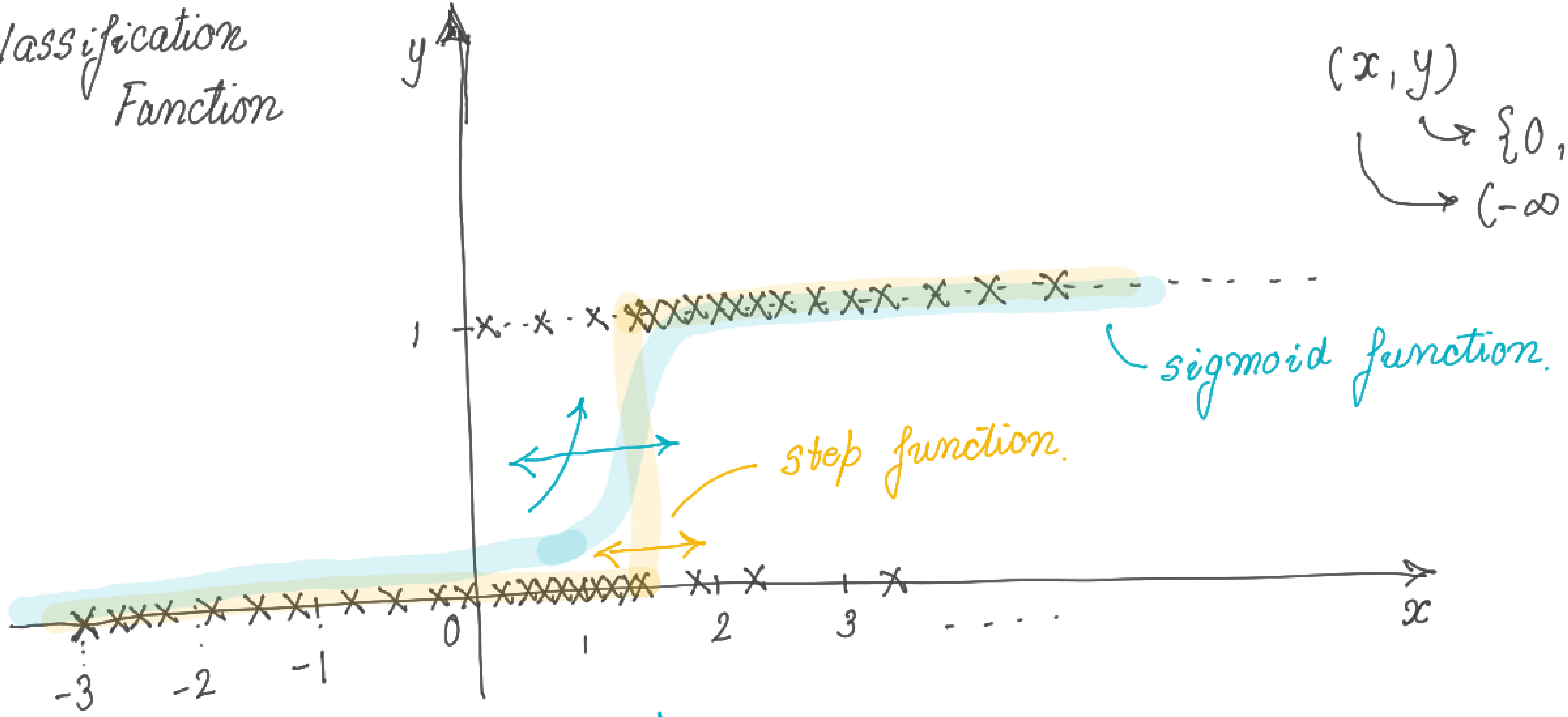
- Ease of computation.
- Transformation from $[0, 1]$ $\downarrow$ $(-\infty, \infty)$

Classification Function

(x, y)

$y \rightarrow \{0, 1\}$

$x \rightarrow (-\infty, \infty)$



$$\sigma(x) = \frac{1}{1 + e^{-\underbrace{kx}_{\text{(shape / steepness)}} + \underbrace{\alpha}_{\text{intercept (location)}}}}$$

Binary Probability Distributi

- Bernoulli
  - Two outcomes $\{0, 1\}$
  - Probability assoc. with each outcome $\begin{cases} 0 & 1-p \\ 1 & p \end{cases}$

$$\boxed{P(X=x) = p^{x}(1+p)^{1-x}} \longrightarrow \text{PMF}$$

$$P(X=0) = p^{0}(1-p)^{1-0} = 1 \times (1-p) = 1-p$$

$$P(X=1) = p^{1}(1-p)^{1-1} = p \times 1 = p$$

$$\underbrace{X1 \quad X2 \quad X3 \;\ldots\; Xp}_{\text{Ind. Var}}$$

$$\underbrace{y}_{\text{Dep. Var}} \rightarrow \{0,1\}$$

$$\hookrightarrow P(y=1) = \underline{p_y} \rightarrow \{0,1\}$$

$f(X_1, X_2 \ldots X_p) \rightarrow y$

?

log odds $(P(y=1))$

f was linear Regression, $\underbrace{y \rightarrow (-\infty, \infty)}$ instead of $\{0,1\}$

$y_i = \sum w_i x_i \rightarrow$ Linear Regression

$\log \text{odds}((y_i = 1)) = \sum w_i x_i \rightarrow$ Logistic Regression

Logistic Regression

Loss f^n: $-\sum \left(y_i \ln(p_i) + (1-y_i)\ln(1-p_i) \right)$ $= L$

$p_i = P(X_i = 1)$

$1 - p_i = P(X_i = 0)$

w_j were found by minimizing LF wrt w_j

(i) $\frac{\partial L}{\partial w_j} = 0 \ \forall j$ → analytically.

(ii) using gradient descent.

$$w_{j,new} \rightarrow w_{j,old} - \underbrace{\eta}_{\text{Learning rate}} \frac{\partial L}{\partial w_{j,old}} \quad \forall w_j$$

$$p_i = \frac{1}{1 - e^{-\sum w_j x_{ij}}}$$

c1	c1_x	c1_y
x	1	0
x	1	0
y	0	1
y	0	1
z	0	0
z	0	0
x	1	0
y	0	1

c1_x, c1_y → dummy variables

0 0 → Z

0 0 → Z.

3 unique vals → 2 columns

Encoding Dummy Variables

n unique values → (n-1) dummy var. cols

c	x_1	x_2	x_3	$y \rightarrow \{0,1\}$
1				0
1				0
1				1
1				0
1				1

sm.add_constant

$$\log\left(\frac{p}{1-p}\right) = w_1 x_1 + w_2 x_2 + w_3 x_3 + w_0 c$$

intercept