Context

dataset: $(X, \underbrace{y}_{})$ → categorical

→ classification

Logistic Regression

$$P(y=1 \mid X) = \frac{1}{1+e^{-Xw}}$$ → parameters

probability of X belonging to class $y=1$ - - - → assumed to follow a functional form (sigmoid function)

For a given X:

$P(y=1 \mid X) = 0.37$

$P(y=0 \mid X) = 0.63$

my value of y should be 0

Context

In Log-Reg: we constrain $P(y=1|X)$ to follow sigmoid function

But what if we calculate $P(y=1|X)$ by some other way?

→ using Bayes' theorem

posterior, likelihood, prior

$$\underbrace{P(y=1|X)}_{\text{posterior}} = \frac{\overbrace{P(X|y=1)}^{\text{likelihood}} \; \overbrace{P(y=1)}^{\text{prior}}}{P(X)} \rightarrow (1)$$

$X = (0,1,0) \qquad y = ? \qquad \Rightarrow P(y=1|X) \;\&\; P(y=0|X)$

$P(y=1) = 2/12 = 1/6$

$P(X|y=1) = 0$

$P(y=1|X) = 0$

$P(y=0|X) = 1$

index	x1	x2	x3	y
1	0	0	1	0
2	1	0	1	0
3	0	1	0	0
4	0	1	0	0
5	0	0	0	0
6	1	0	1	0
7	0	1	0	0
8	0	0	1	1
9	1	0	0	0
10	0	0	1	1
11	0	1	0	0
12	0	1	0	0

$$P(X \mid y=1) = 0$$

$$X \equiv (0,1,0)$$

Assumption

↳ all features are independent of one another

$$P(X \mid y=1) = P(x_1 \mid y=1) \times P(x_2 \mid y=1) \times P(x_3 \mid y=1)$$

Under feature independence;

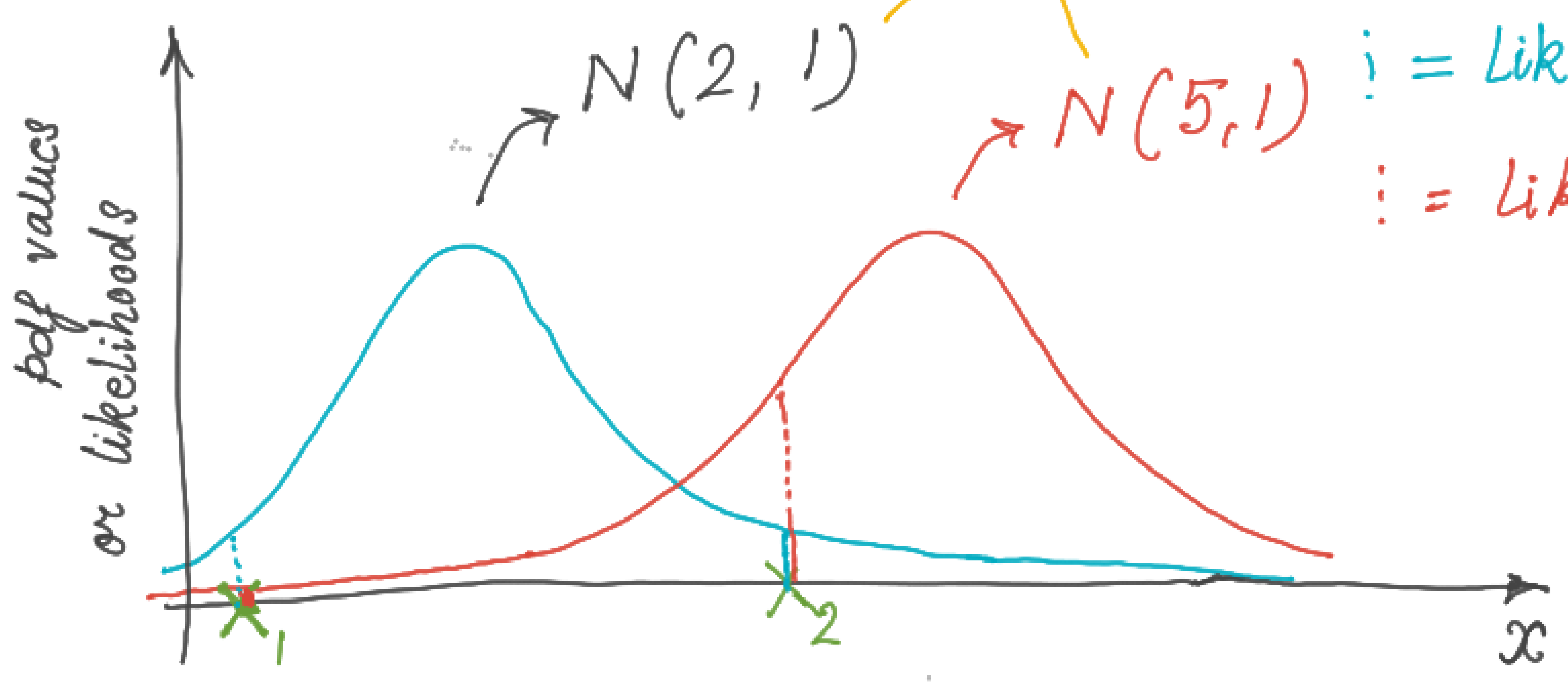
$$P(X \mid y=1) = P(x_1=0 \mid y=1) \times \quad \rightarrow 1$$

$$P(x_2=1 \mid y=1) \times \quad \rightarrow 0 \dashrightarrow \epsilon_2$$

$$P(x_3=0 \mid y=1) \quad \rightarrow 0 \dashrightarrow \epsilon_1$$

Calculating Likelihoods

assumed



= Likelihood of X belonging to blue

= Likelihood of X belonging to red.

$x_1 \rightarrow$ blue

$x_2 \rightarrow$ red.

$$P(\underbrace{X}_{0} \mid \underbrace{y=1}_{\text{blue}}) = pdf(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{0-2}{1}\right)^2} = 0.30$$

$$P(\underbrace{X}_{0} \mid \underbrace{y=0}_{\text{red}}) = pdf(x=0) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{0-5}{1}\right)^2} = 0.0000083$$

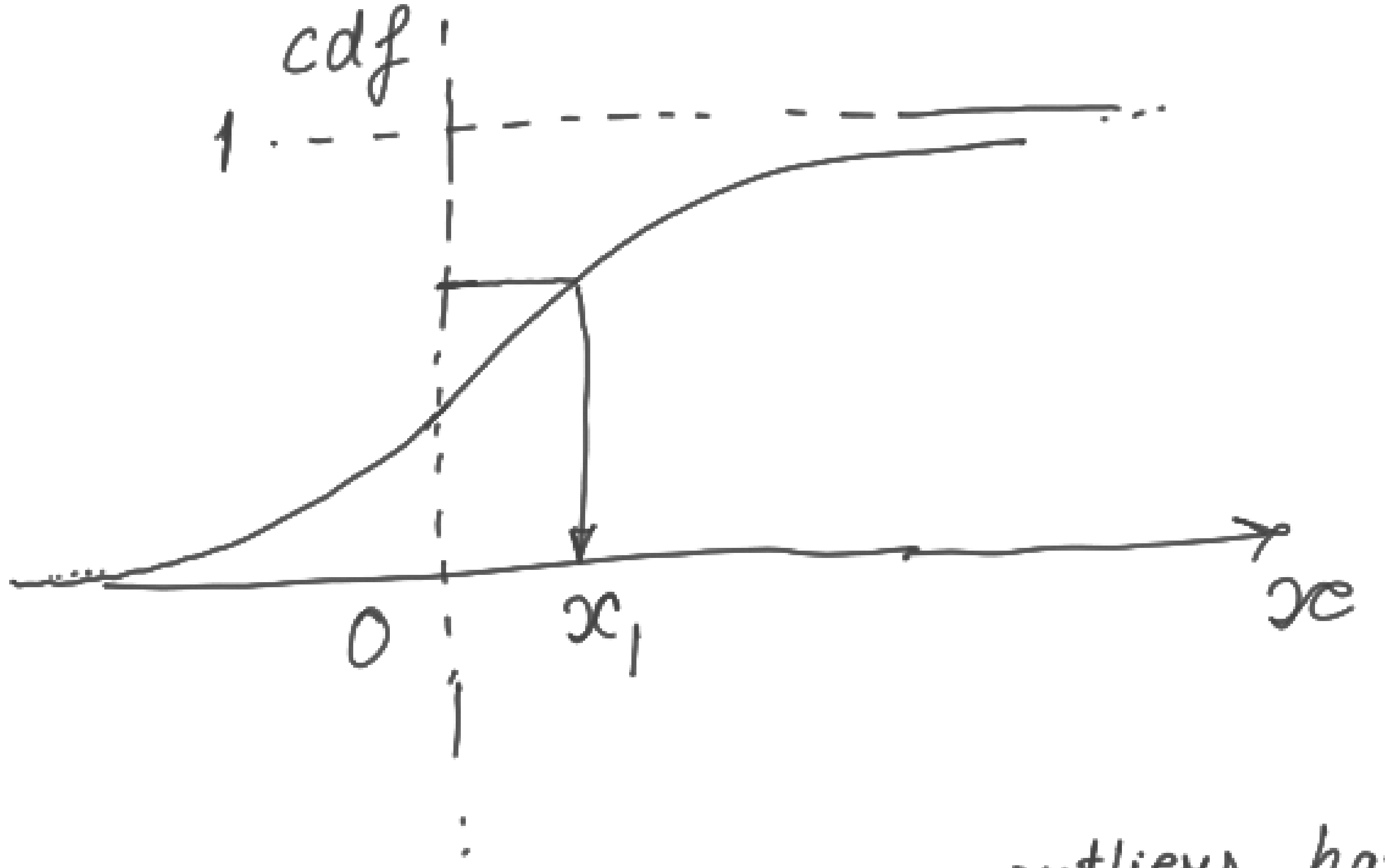
- $(X, y) \rightarrow P(y=1|X)$
- $P(y=1|X)$ is constrained
- model parameters estimated from data
- can't be used to generate new data
- only gives prediction
- logistic regression

DISCRIMINATIVE

- assume $P(X|y=1) \rightarrow$ calculate dist. parameters $\rightarrow$ calc. $P(y=1|X)$
- $P(y=1|X)$ is NOT constrained; but calculated.
- no parameter estimation.
- can be used to generate data
- predicts as well as generates samples
  - Linear/quadratic disc. analysis
  - naive bayes classifier

GENERATIVE

Sampling



$$\left.\begin{aligned} y &\rightarrow U(0,1) \\ x &= F^{-1}(y) \end{aligned}\right\} \text{Sampling using Inverse CDF}$$

F^{-1} → inverse of CDF

$(X, y) \rightarrow$ model $\rightarrow$ parameters $\rightarrow$ prediction : DISCRIMINATIVE

model → outliers handled

$(X, y) \rightarrow$ probability dist. parameters $\rightarrow$ prediction : GENERATIVE