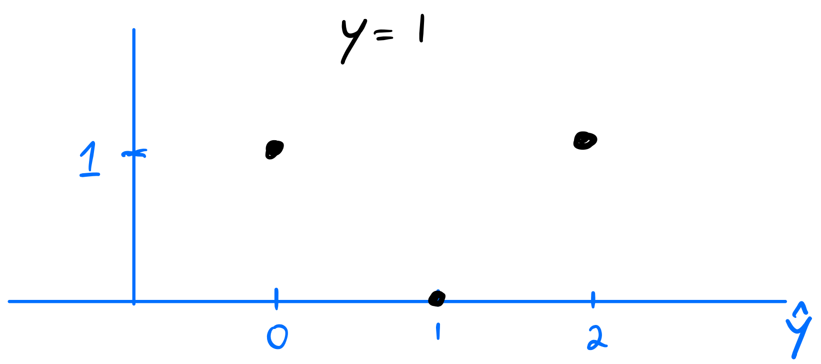
Classification

Labels are unordered (and usually finite)

Ex: Types of wine $\mathcal{Y} = \{\text{Barolo}, \text{Grignolino}, \text{Barbera}\}$
$= \{0, 1, 2\}$

Loss ℓ is 0-1 loss

$$\ell(\hat{y}, y) = \begin{cases} 0 & \text{if } \hat{y} = y \\ 1 & \text{otherwise} \end{cases}$$



$$\hat{L}(f) = \frac{1}{n} \sum_{i=1}^{n} \ell(f(\vec{x}_i), y_i)$$

$$f_{Bayes} = \operatorname*{argmin}_{f \in \{f \mid f: \mathcal{X} \to \mathcal{Y}\}} L(f)$$

$$L(f) = \mathbb{E}\left[\ell(f(\vec{X}), Y)\right]$$

Binary Classification $\mathcal{Y} = \{0, 1\}$

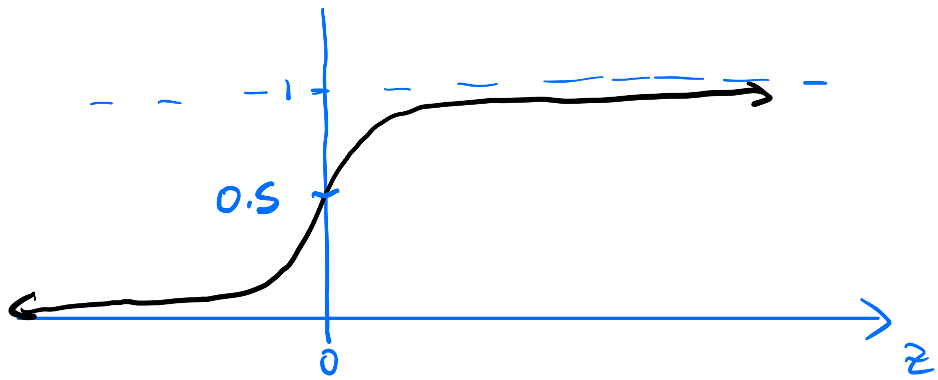
MLE to estimate pmf $p(y|\vec{x})$

$D = ((\vec{X}_1, Y_1), \dots, (\vec{X}_n, Y_n))$

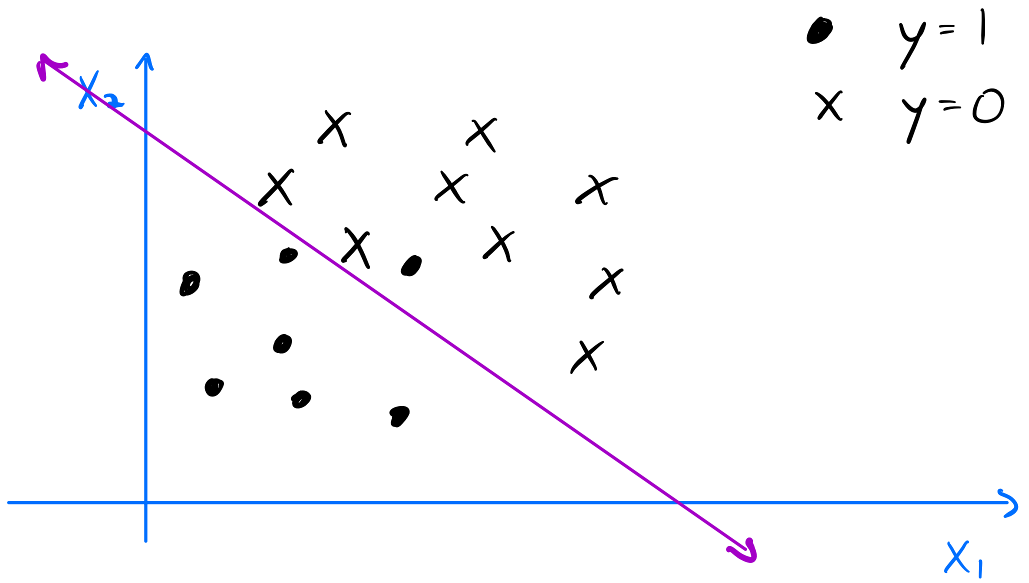
$(\vec{X}_i, Y_i)$ are i.i.d. with $\mathbb{P}_{\vec{X},Y}$, $p_{\vec{x},Y}$

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

"sigmoid" or "logistic"



Ex: $d = 2$



Why not just learn $p(y=1|\vec{x})$ using linear regression?

Ex:

$d=1$