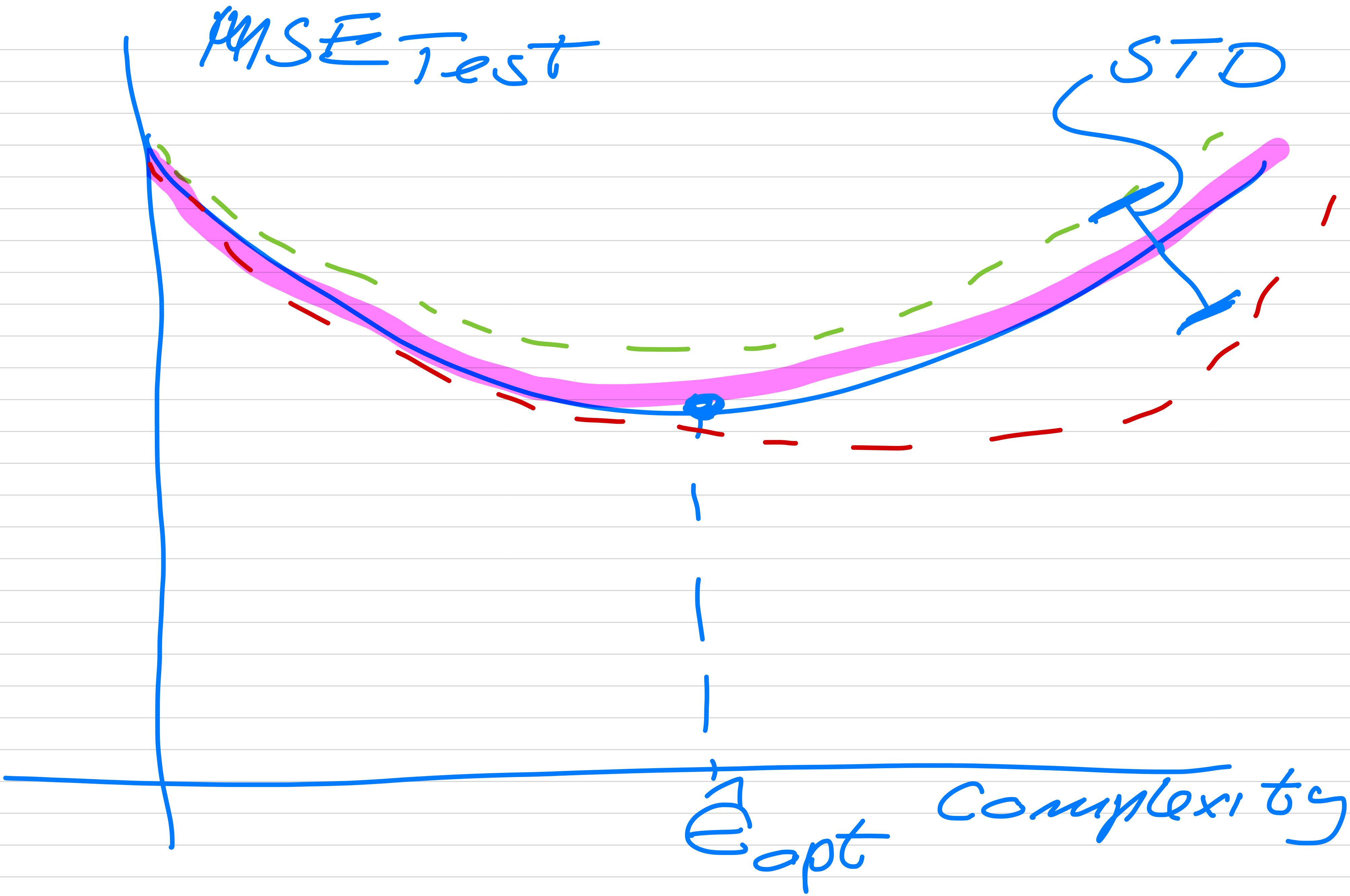
FYS-STK3155/4155, lecture September 22, 2025

FYS-STK3155/4155 September 22

Resampling methods
- Bootstrap
- cross-validation

Classification problems
- Logistic regression


RMSE Test
STD
c_{opt}
Complexity

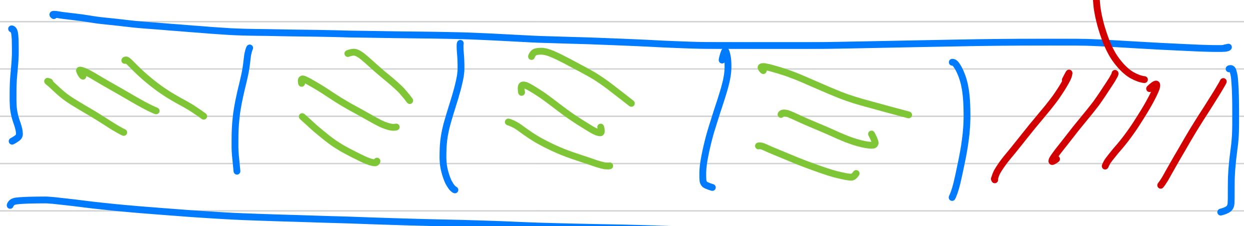
Cross-validation for resampling

- kfold = number of folds

kfold ~ 5-20

kfold = 5

Test



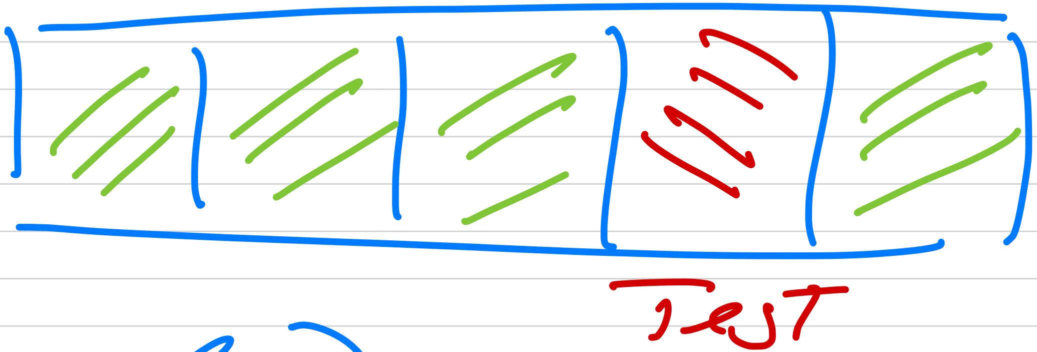
$i = 1$

Training 20%

$MSE_{Test}(1)$

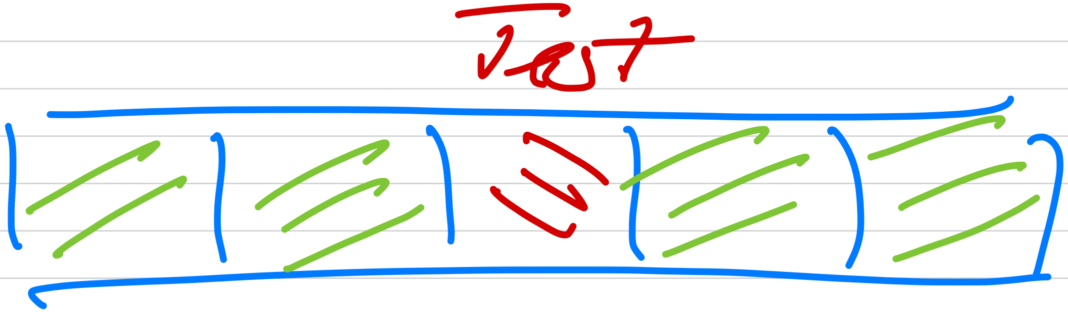
LOOCV (only one in test set)

$i = 2$



$MSE_{Test}(2)$

$i = 3$

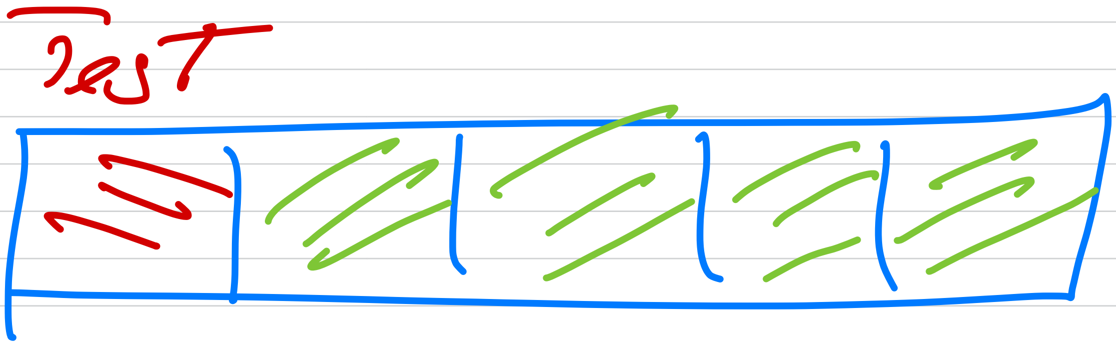


$MSE_{Test}(3)$

$\vdots$

$MSE_{Test}(4)$

$i = 5$



$MSE_{Test}(5)$

$$\overline{MSE_{TEST}} = \frac{1}{5} \sum_{k=1}^{5} MSE_{TEST}(k)$$

Logistic regression

linear regression
continuous

$$y_i = f(x_i) + \varepsilon_i$$

$$\varepsilon_i \sim N(0, \sigma^2)$$

$x_i \in (-\infty, \infty)$ $([x_a, x_b])$

$y_i \in (-\infty, \infty)$ $([y_a, y_b])$

$$y_i \cong \underbrace{\sum_j x_{ij}\,\theta_j}_{\tilde{y}_i = f(x_i)} + \varepsilon_i$$

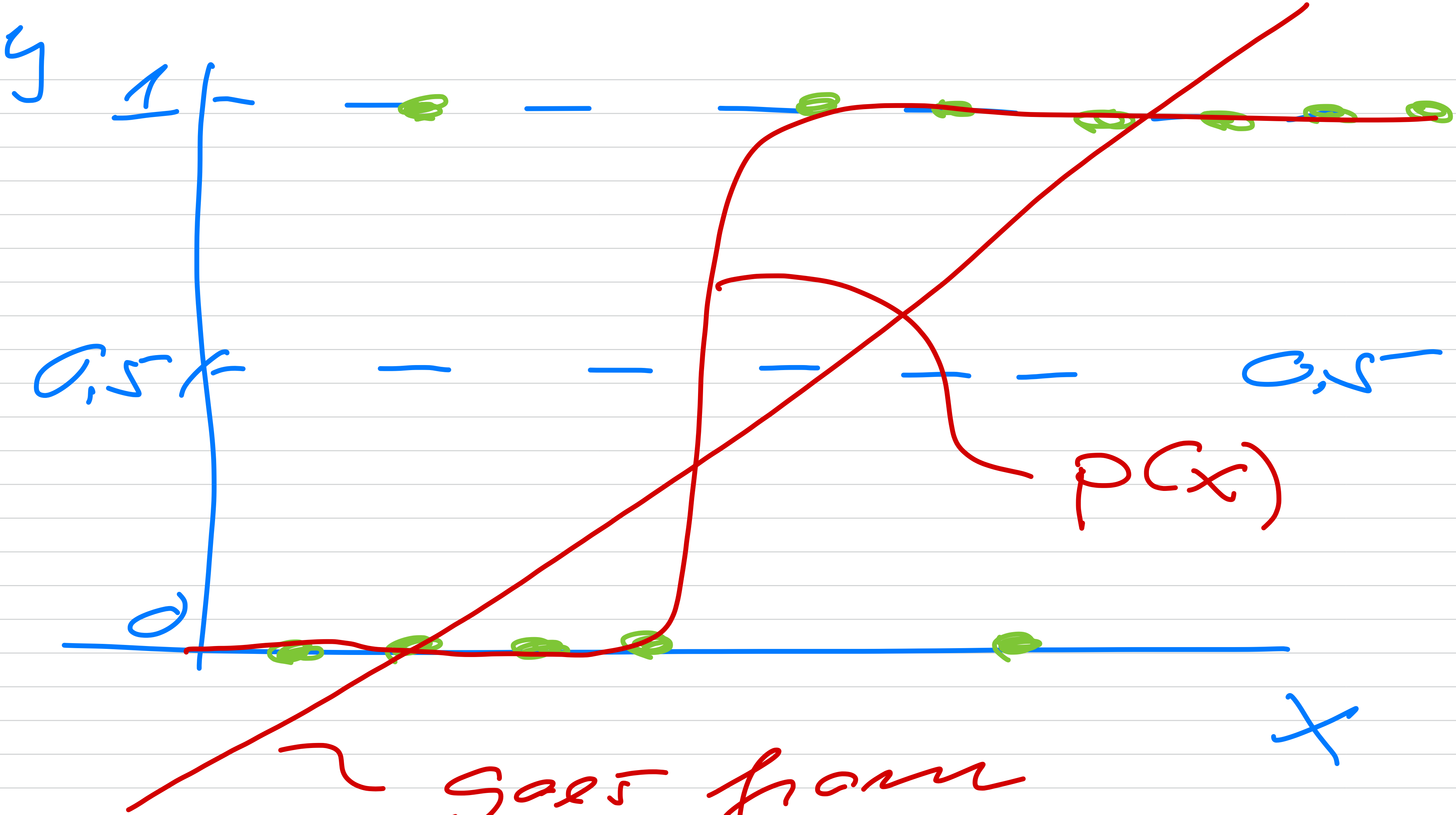
First order polynomial

$$y_i = \theta_0 + \theta_1 x_i + \varepsilon_i$$

what if y_i are discrete?

Example $\quad y_i = \{0, 1\}$

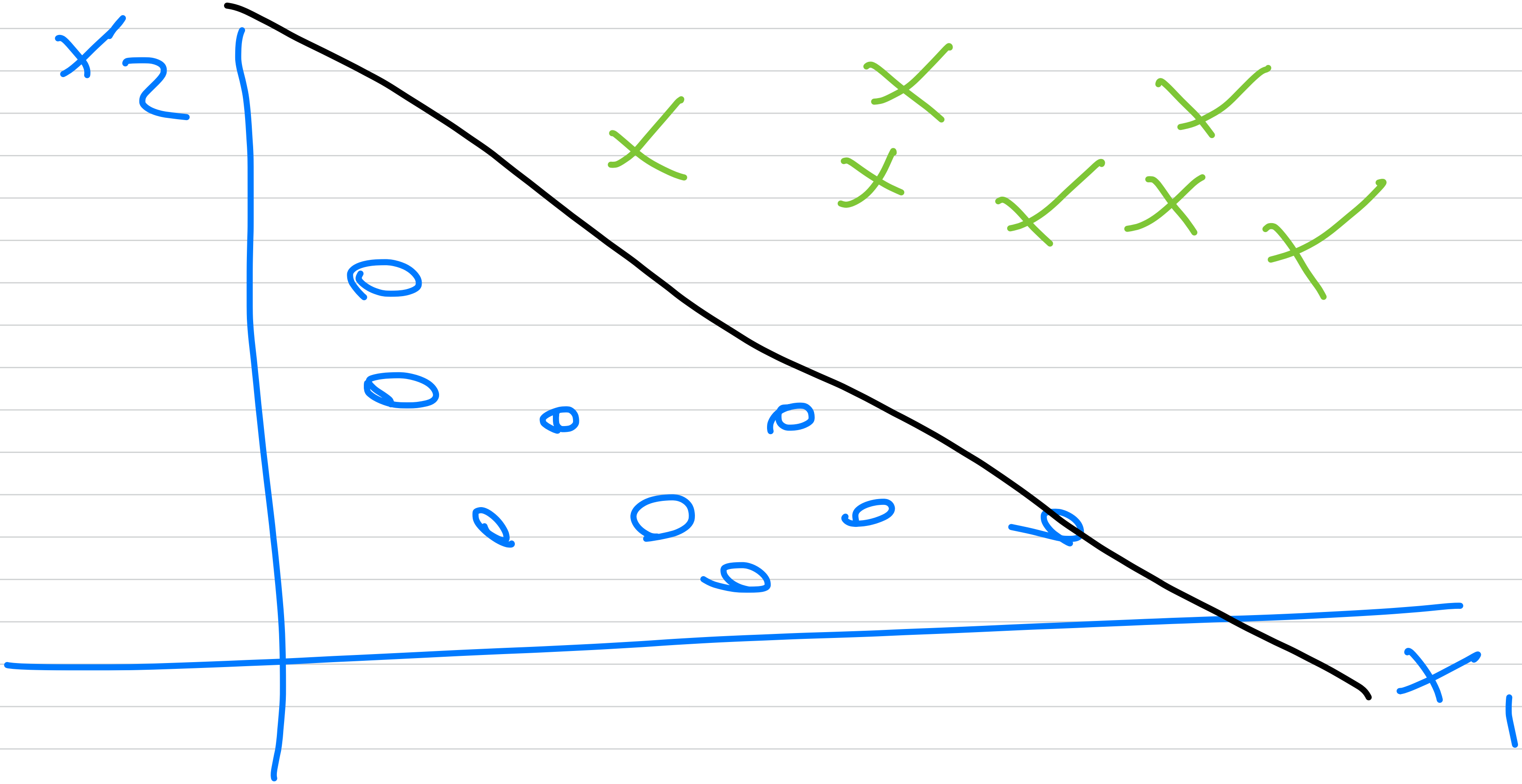
Binary case



goes from $y = -\infty$ to $y = +\infty$

if $p(x_i) \leq 0{,}5$, then $y_i = 0$

else $p(x_i) > 0{,}5$ then $y_i = 1$



$$y_i = f(x_i) + \varepsilon_i$$ replace

$$y_i = p(x_i) + \varepsilon_i$$ can be interpreted as a probability

$$0 \le p(x_i) \le 1$$

$$\int_{x \in D} p(x)\, dx = 1$$

$$\left(\sum_{i \in D} p(x_i) = 1 \right)$$

$$x \in [0, 1] \; ; \; p(1) = 1$$

$$p(x_i) \leq p(x_j) \text{ if } x_i \leq x_j$$

Example: Sigmoid function

$$p(x) = \frac{e^{\theta x}}{1 + e^{\theta x}}$$

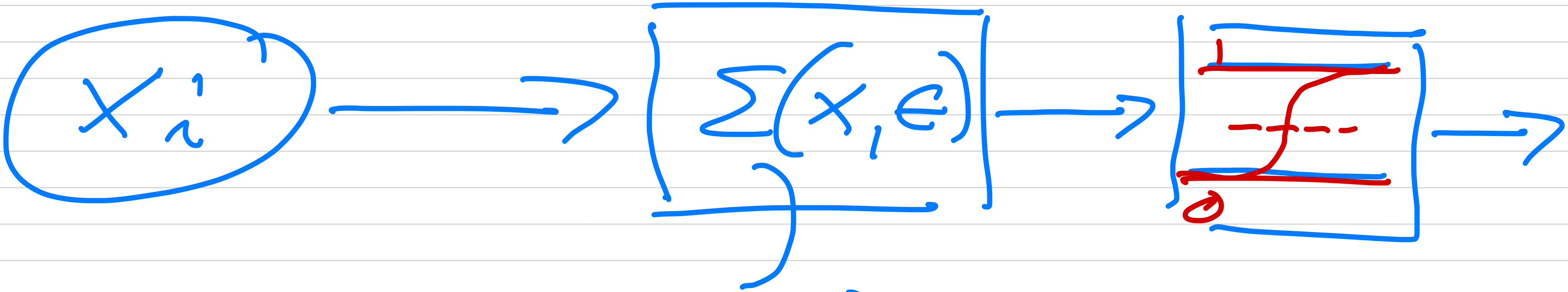
$$y = 1 \; ; \quad p(x) = \frac{e^{\theta x}}{1 + e^{\theta x}}$$

$$y = 0 \; ; \quad 1 - p(x) = \frac{1}{1 + e^{\theta x}}$$

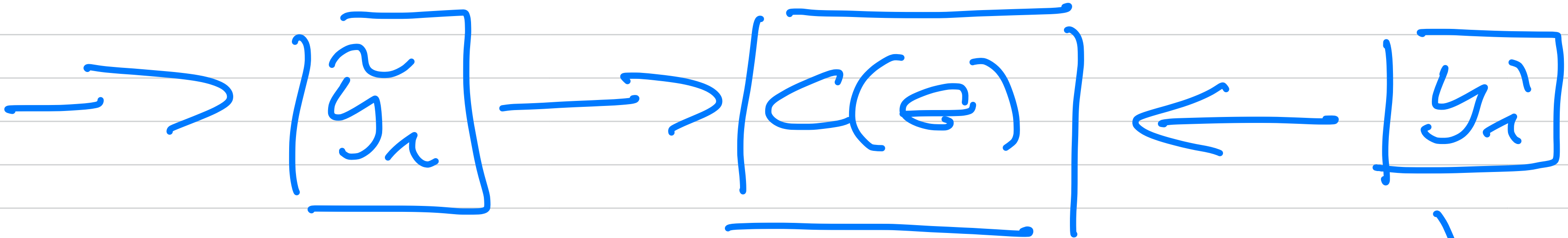
$$\sum_i p(x_i) = p + (1 - p)$$

$$= 1$$

Graphical representation



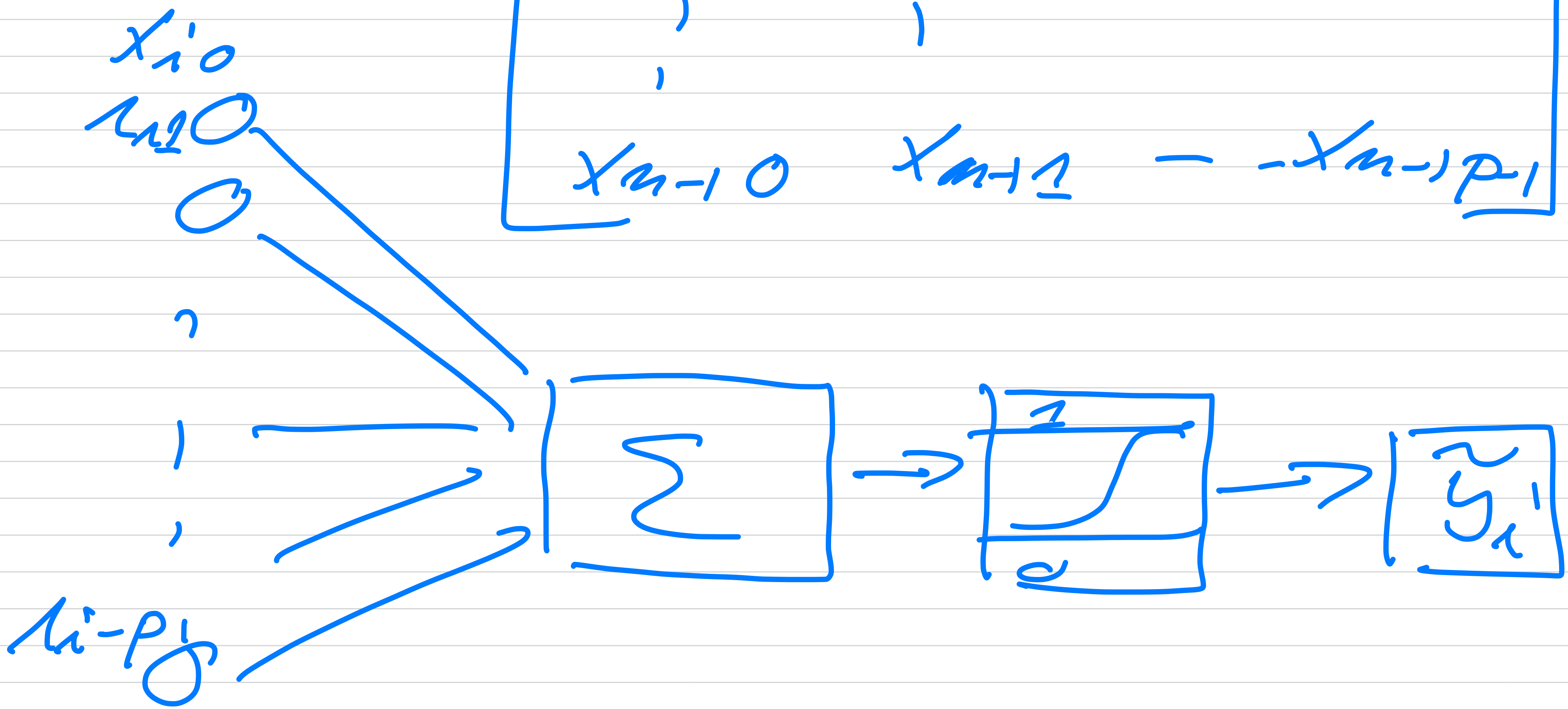
$p(x_i) = p_i$



$\tilde{y}_i = \{0, 1\}$

$y_i = \{0, 1\}$

$$X = \begin{bmatrix} x_{00} & x_{01} & \cdots & x_{0p-1} \\ x_{10} & \vdots & & \\ x_{20} & \vdots & & \\ \vdots & \vdots & & \\ x_{m-10} & x_{m-11} & \cdots & x_{m-1p-1} \end{bmatrix}$$



Model

$$p(x_i) = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}$$

$$D = \{(x_0, y_0), (x_1, y_1), \ldots, (x_{n-1}, y_{n-1})\}$$

$$y_i = p(x_i) + \varepsilon_i = p_i + \varepsilon_i$$

$$P(D \mid \theta) = c(\theta)$$

$$P(D \mid \theta) = \prod_{i=0}^{n-1} P_i^{y_i} (1-P_i)^{1-y_i}$$

$$y_i = \underline{1} \; ; \; P_i$$

$$y_i = 0 \; ; \; 1-P_i$$

$$\frac{\partial P(D \mid \theta)}{\partial \theta_j} = 0$$

$$C(\theta) = -\log P(D|\theta)$$

$$\hat{\theta} = \underset{\theta \in \mathbb{R}^p}{\arg\min} \left(-\log P(D|\theta)\right)$$

$$\prod_{i=0}^{n-1} p_i^{y_i} (1-p_i)^{1-y_i}$$

$$C(\theta) = -\sum_{i=0}^{m-1}\Big[y_i' \log p_i' + (1-y_i')\log(1-p_i')\Big]$$

$$\log\left[\frac{e^{\theta_0+\theta_1 x_i'}}{1+e^{\theta_0+\theta_1 x_i'}}\right]$$

$$= \left(\theta_0 + \theta_1 x_i'\right)$$

$$- \log\left(1+e^{\theta_0+\theta_1 x_i'}\right)$$

$$C(\theta) =$$

$$- \sum_{i=0}^{m-1} \left[y_i \left(\theta_0 + \theta_1 x_i \right) - \log\left(1 + e^{\theta_0 + \theta_1 x_i}\right) \right]$$

$$\frac{\partial C}{\partial \theta_0} = 0 \quad \wedge \quad \frac{\partial C}{\partial \theta_1} = 0$$

$$\frac{\partial C}{\partial \theta_0} = 0 = - \sum_{i=0}^{m-1} \left(y_i - p_i \right)$$

$$\frac{\partial C}{\partial \theta_1} = 0 = -\sum_{i=0}^{n-1} x_i'(y_i' - p_i)$$

$$\frac{\partial C}{\partial \theta} = -X^T(\vec{y} - \vec{p})$$

depends on θ

$$p(x_i') = \frac{e^{\theta_0 + \theta_1 x_i'}}{1 + e^{\theta_0 + \theta_1 x_i'}}$$

$$\frac{\partial^2 C}{\partial \theta^2} = X^T W X$$

$$w_{ii} = p_i (1 - p_i)$$

$$w_{ij} \ (i \neq j) = 0$$

$$\theta^{(m+1)} = \theta^{(m)} - \eta \left. \frac{\partial C}{\partial \theta} \right|_{\theta = \theta^{(m)}}$$