Classroom of 50 students given a basketball and asked to score $\begin{cases} 0 & \text{NO SCORE} \\ 1 & \text{SCORED} \end{cases}$

H → continuous

S → discrete

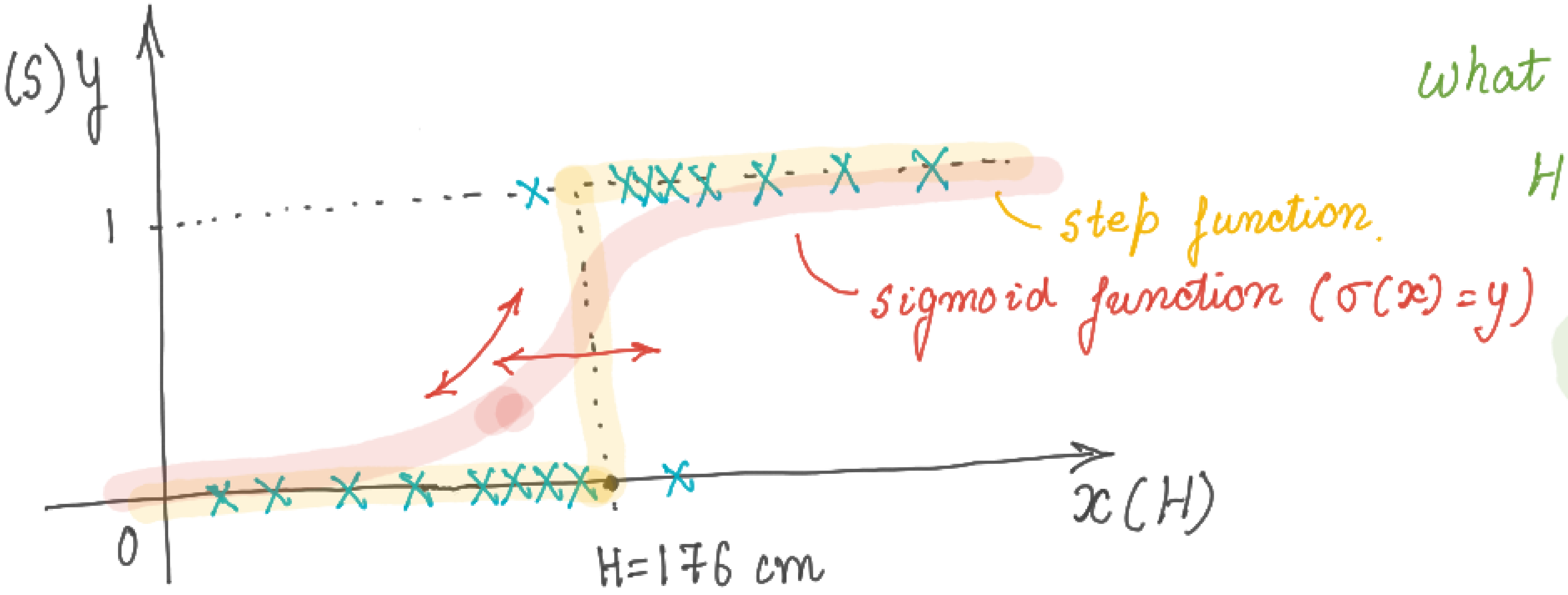
ID	H	S
1	179	1
2	162	0
3	167	0
4	175	1
5	181	1
6	183	0
7	159	0
8	⋮	⋮
⋮	⋮	⋮
13	177	? → probability → [0,1]

Dataset

Can we know from the height of a student; whether they will score or not?

⇓

Can we know the probability of scoring given a person's height?



what is the relationship b/w H and S ?

$$\sigma(H) = \frac{1}{1+e^{-k(H-p)}}$$

$\rightarrow [0, 1]$

The sigmoid function can be used to model probabilities as a function of one/many continuous/categorical variables.

w_0 x_1 x_2 x_3 $y \rightarrow \{0, 1\}$

w_1 w_2 w_3

$$\underbrace{w_0 + w_1 x_1 + w_2 x_2 + w_3 x_3}_{A} \equiv$$ combines 3 ind. var. into one scalar qty

can be passed to sigmoid.

$$\sigma(A) = \frac{1}{1+e^{-A}} = p(\text{event}) = p(y=1) \; (= p)$$

$$\Rightarrow p = \frac{1}{1+e^{-A}} \qquad 1-p = 1 - \frac{1}{1+e^{-A}} = \frac{e^{-A}}{1+e^{-A}}$$

$$\Rightarrow \log_e\left(\frac{p}{1-p}\right) = \log_e\left(\frac{1}{e^{-A}}\right) = A = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_3$$

Consider that you have

3 ind. var : x_1, x_2, x_3

1 dep. var : $y \to \{0, 1\}$

Constraint is ; I want to essentially re-use linear regression.

↳ $y \in (-\infty, \infty)$

✓ I model the relationship between (x_1, x_2, x_3) and $\underbrace{P(y=1)}$

↳ $[0,1] \longrightarrow [0, \infty)$

$$\frac{P(y=1)}{1-P(y=1)} = \text{Odds}(P(y=1))$$

$[0, \infty)$

$\downarrow$

$(-\infty, \infty)$

$$\text{Log}(\text{Odds}(P(y=1))) = \ln\left(\frac{P(y=1)}{1-P(y=1)}\right)$$

$$\underbrace{\log \text{Odds}\,(P(y=1))}_{\text{Transformed target variable.}} = \underbrace{w_0 + w_1 x_1 + w_2 x_2 + w_3 x_3}_{\text{Linear terms.}}$$

Binary Probability Distribution → Bernoulli Dist

↳ models outcomes of exp with 2 outcomes

$$\text{Tossing a coin} \rightarrow \left\{ \begin{array}{ccc} H & 0 & 1-p \\ T & 1 & p \\ & X & \end{array} \right\} \text{Bern. Dist}^n$$

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

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

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

$x_1 \quad x_2 \quad x_3 \quad \cdots \quad x_p \quad y$

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

$$\log\left(\frac{p}{1-p}\right) = \underbrace{\sum w_i \cancel{X}_i}$$

$$p = P(y=1)$$

$$\text{Loss } f^n = -\sum_{i=1}^{N}\Big[\underbrace{y_i \log p_i + (1-y_i)\log(1-p_i)}_{\text{log of the PMF of Bernoulli Dist.}}\Big]$$

= Binary Cross Entropy Loss

$$p_i = \frac{1}{1+e^{-\sum w_i X_i}}$$

To find w_i;

$$\frac{\partial L}{\partial w_i} = 0 \quad \forall\, i \quad \rightarrow \text{analytical}$$

To find w_i :

1. Assume random w_i

2. Calculate new w_i's as

learning rate.

$$w_{i,new} = w_{i,old} - \eta \frac{\partial L}{\partial w_{i,old}}$$

3. Repeat (2) until convergence criteria.

↳ total # of steps

↳ • $|w_{i,new} - w_{i,old}| < \epsilon$