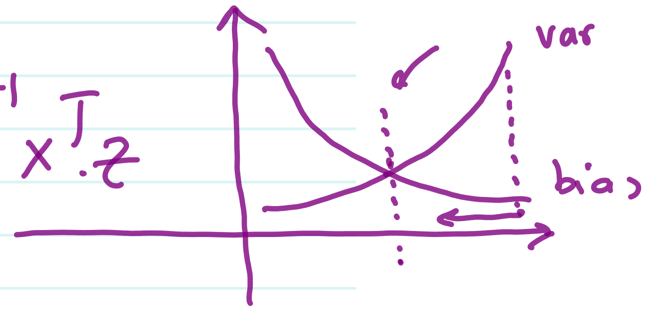
Logistic Regression:

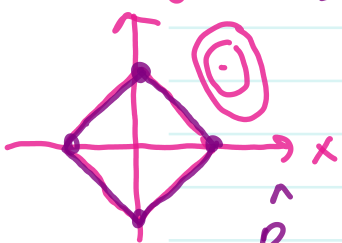
$$\hat{z} = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p \qquad \text{MLR}$$

$$\Downarrow$$

$$SSE = \|z - XB\|_2^2 \rightarrow \hat{\beta}_{LSE} = (X^T X)^{-1} X^T z$$



L1 -

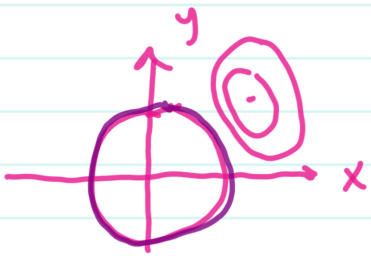


$$\hat{\beta}_{L_1} = \underset{\beta}{\mathrm{Arg\,min}} \left[\|z - XB\|_2^2 + \lambda \|B\|_1 \right]$$

$$\|B\|_1 = \sum_{i=1}^{p} |\beta_i|$$

L2 -

$$\hat{B}_{L_2} = \mathrm{Arg\ min} \left[\|z - XB\|_2^2 + \lambda \|B\|_2^2 \right] \qquad \|B\|_2 = \sqrt{\sum_{i=1}^{p} B_i^2}$$



when to use L_1 : for feature reduction
when to use L_2 : for prediction

L_1 followed by L_2

Elastic Net

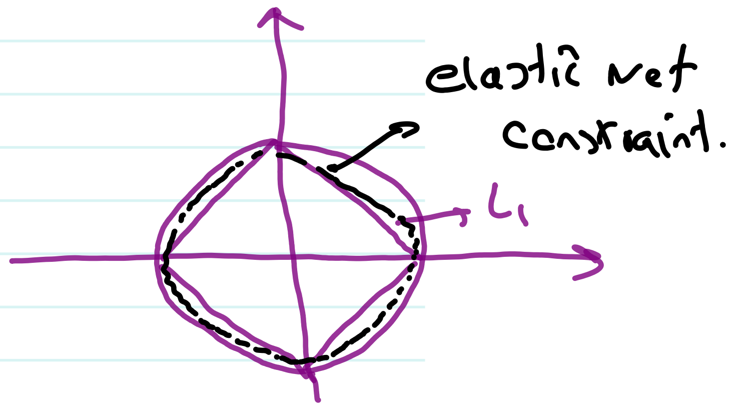
$$\hat{\beta}_{elastic} = \underset{\beta}{\text{Arg min}} \left[\|z - X\beta\|_2^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2 \right]$$

$$= \underset{\beta}{\text{Arg min}} \left[\|z - X\beta\|_2^2 + \alpha \rho \|\beta\|_1 + \alpha (1-\rho) \|\beta\|_2 \right]$$

$\alpha = 0 \longrightarrow$ LES – No Reg.

$\alpha = 1 \longrightarrow \rho = 1 \longrightarrow L_1$ reg.

$\alpha = 1 \longrightarrow \rho = 0 \longrightarrow L_2$ reg



$\alpha = 1 \to 0 < p < 1 \to$ Elastic net

L2 reg $\to$ $\boxed{\hat{\beta}_{L_2} = (X^T \cdot X + \lambda I)^{-1} X^T \cdot z}$

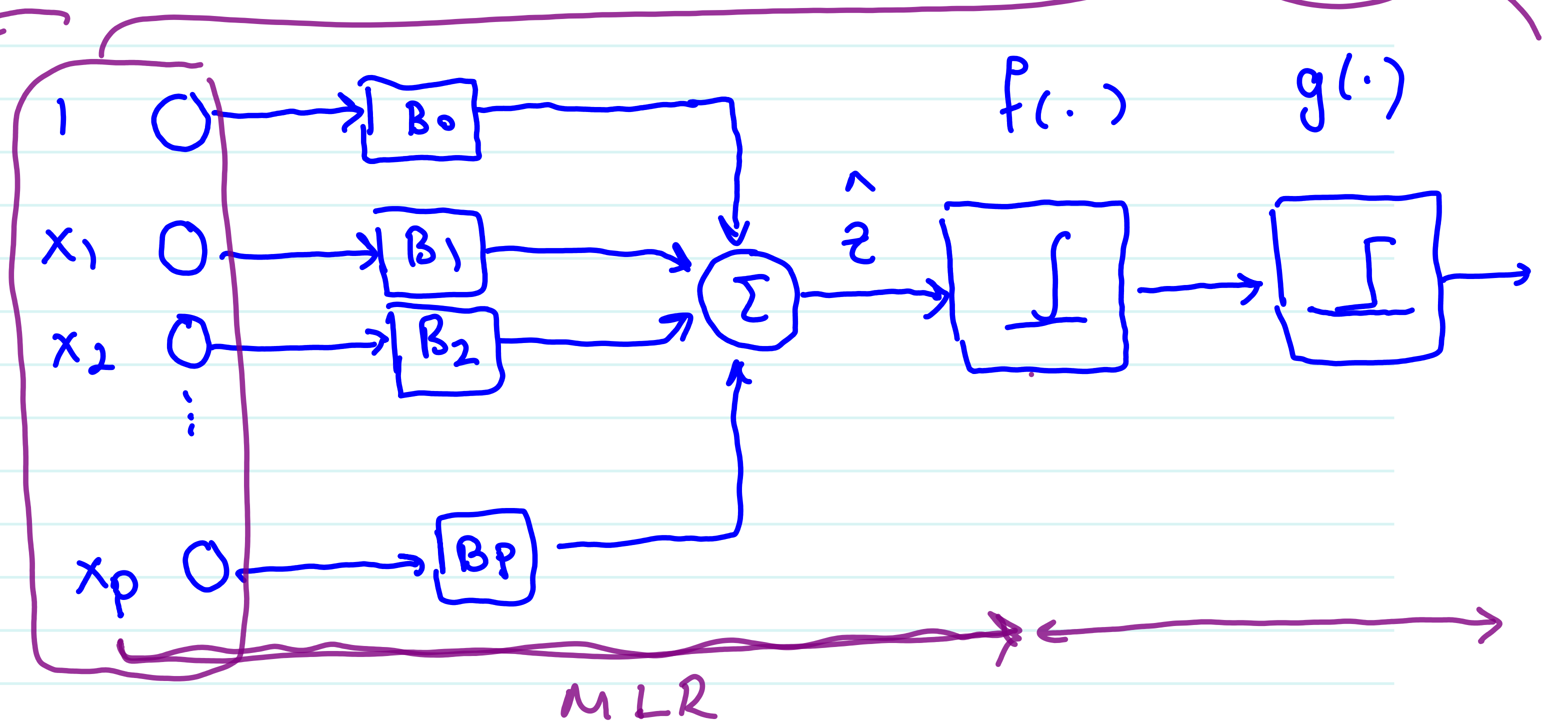
$$\hat{z} = \overbrace{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}$$ MLR

logistic regressio

Neural Network

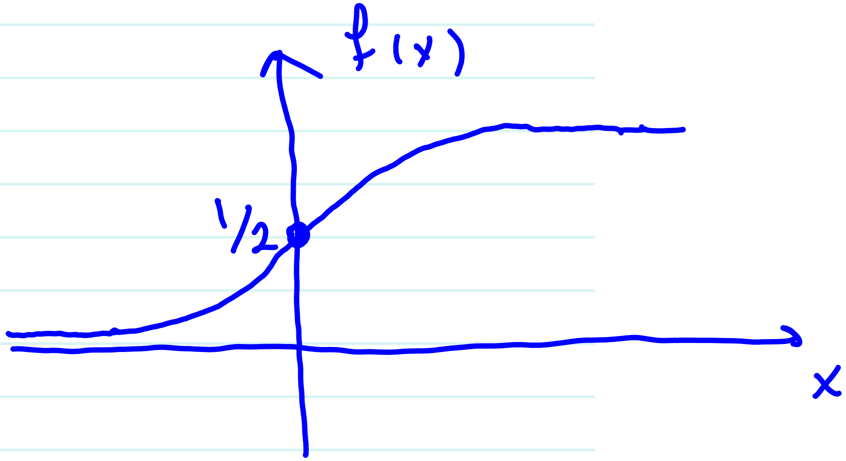
1 layer

1 neuron.



$$g(x) = \begin{cases} 1 & : \quad x \geq 0.5 \\ 0 & : \quad \text{Else} \end{cases}$$

$$\boxed{f(x) = \frac{1}{1+e^{-x}}}$$



$$\lim_{x \to -\infty} f(x) = 0$$

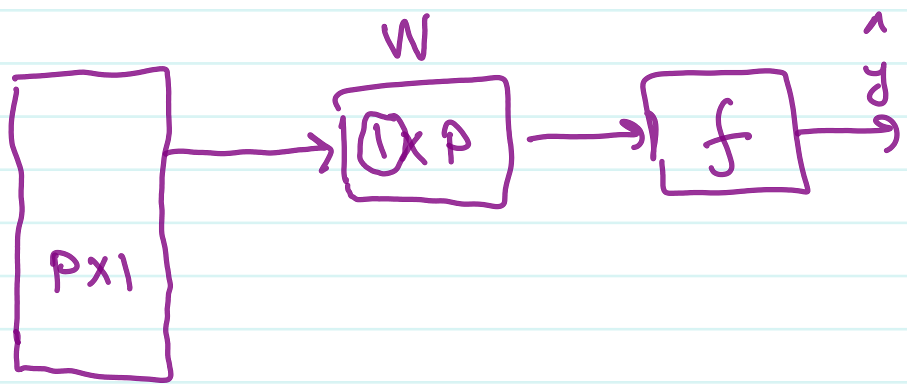
$$\lim_{x \to +\infty} f(x) = 1$$

$$f(0) = \frac{1}{1+e^{0}} = \frac{1}{2}$$

Logisti

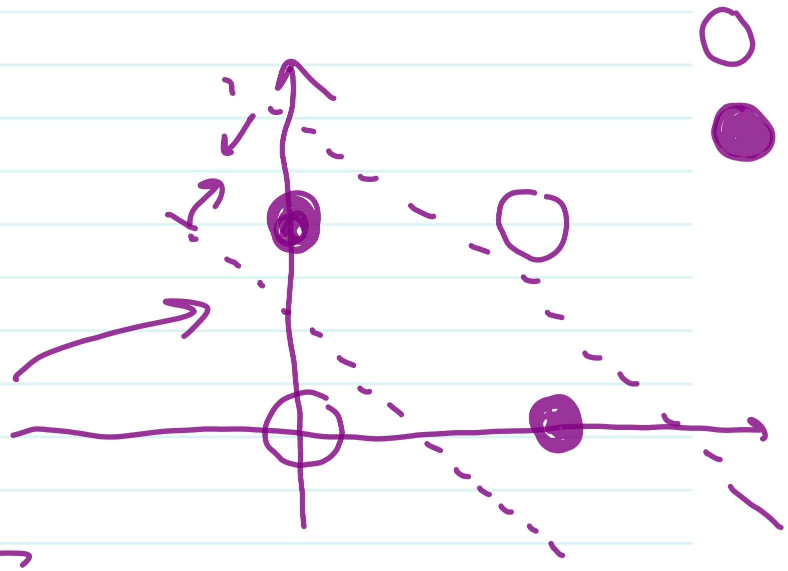
(∞, ∞) $\xrightarrow[f(\cdot)]{\text{regression}}$ $[0, 1]$ → similar to probability.

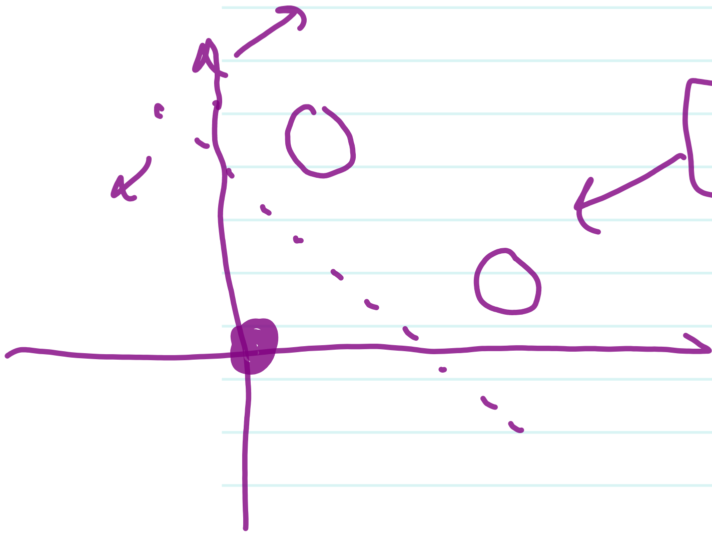
Block diagram of Logistic Regression



Linearly Seprable :

Not linearly seprable

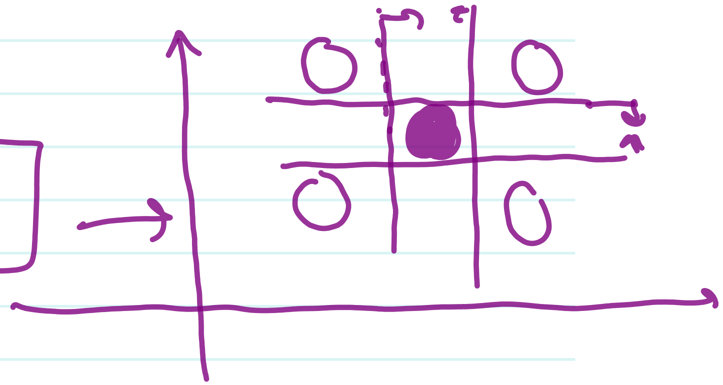




Linear seprable

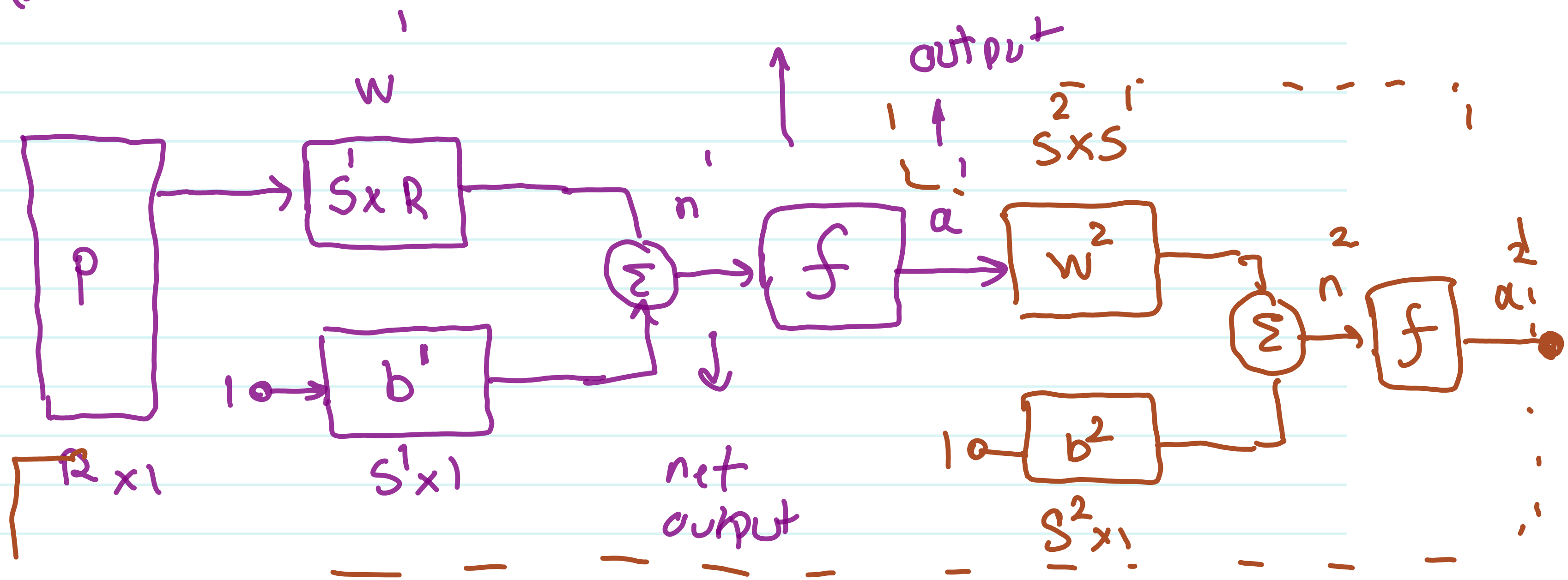
or

Not linearly seprable



MLP

Muti precptron



$$n^1 = W^1 P + b^1 \qquad n^i = W^i a^{i-1} + b^i$$

$$a^1 = f(n^1) \qquad a^i = f(n^i)$$

a 2 layerd N.N with sufficient # of neurons

is called $\boxed{\text{Universal Approximator}}$

$\boxed{\text{NARX}}$ Nonlinear Autoregressive with exogenous input

---

$$\frac{P(1 \mid z)}{P(0 \mid z)} = e^{z} \rightarrow \frac{P(1 \mid z)}{1 - P(1 \mid z)} = e^{z}$$

$$P(1 \mid z) = e^{z} - e^{z} P(1 \mid z)$$

$$P(1 \mid z)\left(1 + e^{z}\right) = e^{z}$$

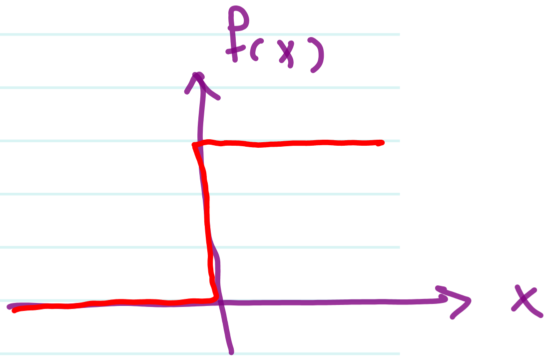
$$P(1 \mid z) = \frac{e^{z}}{1+e^{z}} = \frac{1}{\frac{1}{e^{z}} + 1} = \boxed{\frac{1}{1+e^{-z}}}$$

· logit $\longrightarrow$ ln of odd function

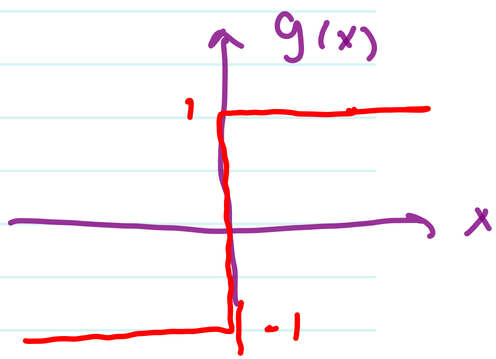
$$\ln\left(\frac{p(1|z)}{p(0|z)}\right) = \ln(e^{z}) = z$$

$$\boxed{\text{logit}(\text{odd}) = z = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}$$

$$f(x) = \text{hardlim}(x) = \begin{cases} 1 & : x \geq 0 \\ 0 & : \text{Else} \end{cases}$$



$$g(x) = \text{hardlims}(x) = \begin{cases} 1 & : x \geq 0 \\ -1 & : x < 0 \end{cases}$$



Example of logistic regression

target	← variable →	intercept	Coef. x_1 (age)
Subscription to magazine		-26.52	0.78

$$z = B_0 + B_1 x_1$$

$$\boxed{z = -26.52 + 0.78 x_1}$$

So if person age is $\underline{35}$

* what is the probability that the $\underset{x^*}{\underline{\text{person}}}$ subscribe to magazine

$$z^* = -26.52 + 0.78(35) = 0.78$$

$$y^* = \sigma(z^*) = \frac{1}{1+e^{-z^*}} = \frac{1}{1+e^{-0.78}} = 0.68 \text{ or } 68\%$$

What about a person @ age = 36 $\longrightarrow$ $P(z^*) = ?$

$\underline{\Delta P} = P(\text{age } 36) - P(\text{age } 35)$

$$y^* = \sigma(z^*) = \frac{1}{1+e^{-z^*}} = \frac{1}{1+e^{-(-28.52 + 0.78 * 36)}}$$

$$= 82\%$$

$$\Delta P = 82\% - 68\% = 14\%$$

So for a person with 25 years old

$$P(25) = .009 = 0.9\%$$
$$P(26) = 0.002 = 0.2\%$$
$$\Delta P = -0.7\%$$

	coef	35	36	25	26	45	46
Constant	−26.52						
Age	0.78						
P( )		69%	83%	.9%	0.2%	1	1