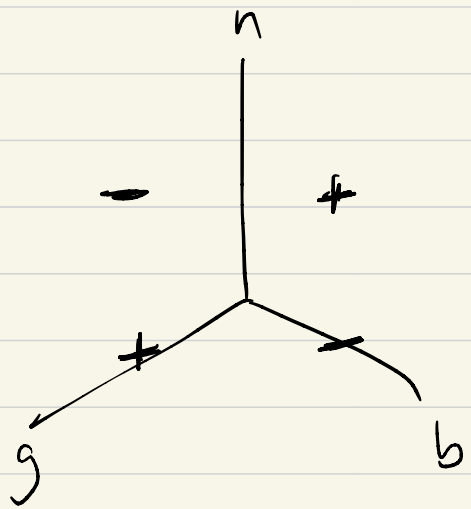
CS 371N Lecture 5: Fairness, Neural Networks

Fairness: see slides

Neural Nets

$f(\bar{x}) = [g \; b \; n]$

good	1	0	0
bad	0	1	0
not good	1	0	1
not bad	0	1	1



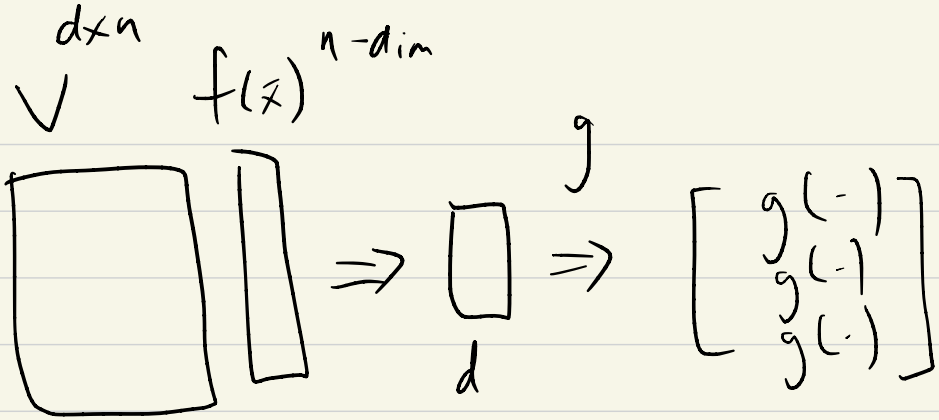
Linear

$\bar{w}^T f(\bar{x})$ ≠ intermediate feats.

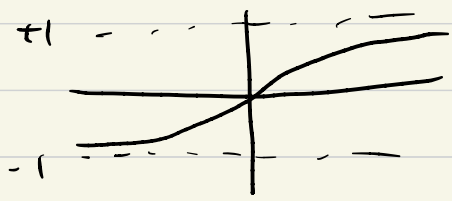
$\bar{z} = g(V f(\bar{x}))$ $f(\bar{x})$ is length n

V ← matrix $d \times n$

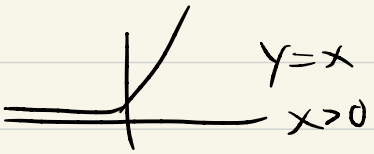
g ↑ elementwise nonlinearity



g: tanh

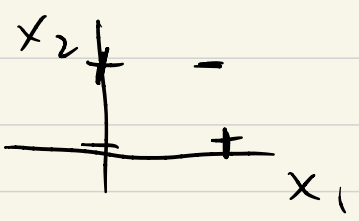


ReLU



y=0
x<0

XOR:



Not linearly separable

x_1	x_2	y
0	0	0
1	0	1
0	1	1
1	1	0

I made this up

$$V = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$$

$g = \tanh$

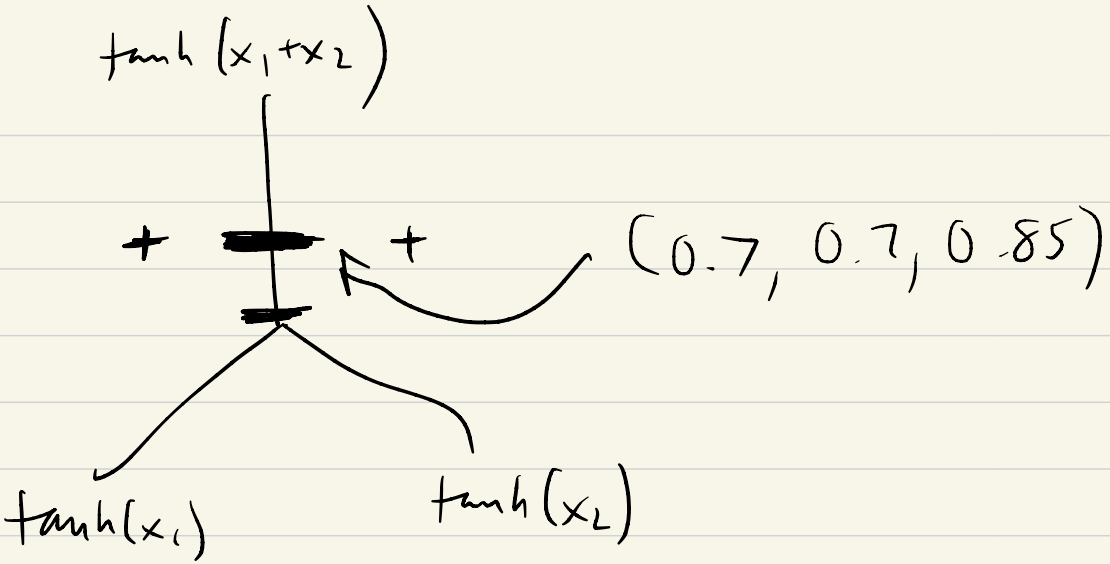
$$Vf(\bar{x}) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_1 + x_2 \end{bmatrix}$$

$$g(Vf(\bar{x})) = \begin{bmatrix} \tanh(x_1) \\ \tanh(x_2) \\ \tanh(x_1 + x_2) \end{bmatrix}$$

$\tanh(0) = 0$

$\tanh(1) = 0.7$

$\tanh(2) = 0.85$



Classifier: $\bar{w}^T \bar{z} = \bar{w}^T g(Vf(\bar{x}))$

data is separable in new space