show decision boundary of naive Bayes is also linear

1)

predict label 1 if $p(t=1 \mid \mathbf{x}) \geq p(t=0 \mid \mathbf{x})$

equivalently, $$\frac{p(\mathbf{x} \mid t=1)\, p(t=1)}{p(\mathbf{x} \mid t=0)\, p(t=0)} \geq 1$$

naive Bayes assumption : $p(\mathbf{x} \mid t) = \prod_{i=0}^{d} p(x_i \mid t)$

$$\frac{p(t=1)}{p(t=0)} \prod_{i=0}^{d} \frac{p(x_i \mid t=1)}{p(x_i \mid t=0)} \geq 1$$

let $p(t=1) = p$

let $p(x_i = 1 \mid t=1) = a_i$

let $p(x_i = 1 \mid t=0) = b_i$

$$\frac{p}{1-p} \prod_{i=0}^{d} \frac{a_i^{x_i} (1-a_i)^{(1-x_i)}}{b_i^{x_i} (1-b_i)^{(1-x_i)}} \geq 1$$

$$\left(\frac{p}{1-p} \prod_{i=0}^{d} \frac{1-a_i}{1-b_i} \right) \prod_{i=0}^{d} \left(\frac{a_i (1-b_i)}{b_i (1-a_i)} \right)^{x_i} \geq 1$$

take log

1)
continued

$$\log\left(\frac{p}{1-p}\prod_{i=0}^{d}\frac{1-a_i}{1-b_i}\right) + \sum_{i=0}^{d} x_i \log\left(\frac{a_i(1-b_i)}{b_i(1-a_i)}\right) \geq 0$$

for any input x, first term is constant because it does not have any x_i terms

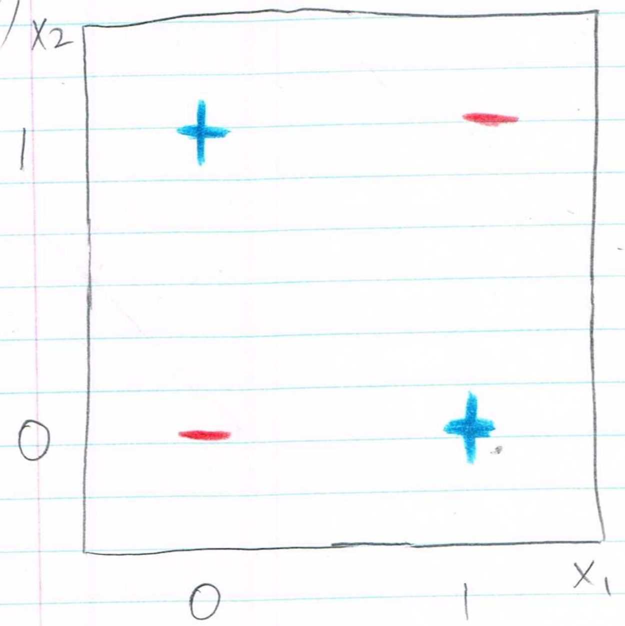
let $b = \log\left(\frac{p}{1-p}\prod_{i=0}^{d}\frac{1-a_i}{1-b_i}\right)$

let $\log\left(\frac{a_i(1-b_i)}{b_i(1-a_i)}\right) = w_i$

$$\left(\sum_{i=0}^{d} x_i w_i\right) + b \geq 0$$

this is linear

2)



3 linear classifiers:

$\phi_1(x) = x_1$

$\phi_2(x) = x_2$

$\phi_3(x) = x_1 x_2$

$\phi(x) = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$

x_1	x_2	$\phi_1(x)$	$\phi_2(x)$	$\phi_3(x)$	t
0	0	0	0	0	0
0	1	0	1	0	1
1	0	1	0	0	1
1	1	1	1	1	0

$b = -0.5 \quad w_1 = 1 \quad w_2 = 1 \quad w_3 = 2, \quad w = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$

These weights & biases work for XOR

Show optimization of neural networks is non convex

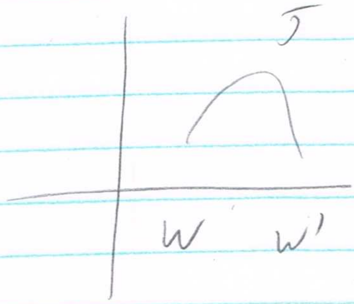
3)

Optimizing neural networks is non-convex

there could be multiple local minima

$J(w)$ small $\quad J(w')$ small

$J(\frac{1}{2}w + \frac{1}{2}w')$ large



Sigmoid neuron

$$Y = \frac{1}{1+e^{-(\Sigma w_i x_i + b)}} = \frac{1}{1+e^{-(w_0 + w_1 c_1 + w_2 c_2)}}$$

rename c_1 as c_2 and c_2 as c_1

$$= \frac{1}{1 + e^{-\left(w_0 + w_1\left(\frac{1}{1+e^{-(w_{10} + w_{11}x_1 + w_{12}x_2)}}\right) + w_2\left(\frac{1}{1+e^{-(w_{20} + w_{21}x_1 + w_{22}x_2)}}\right)\right)}}$$

computation of this gradient is tedious but also non convex