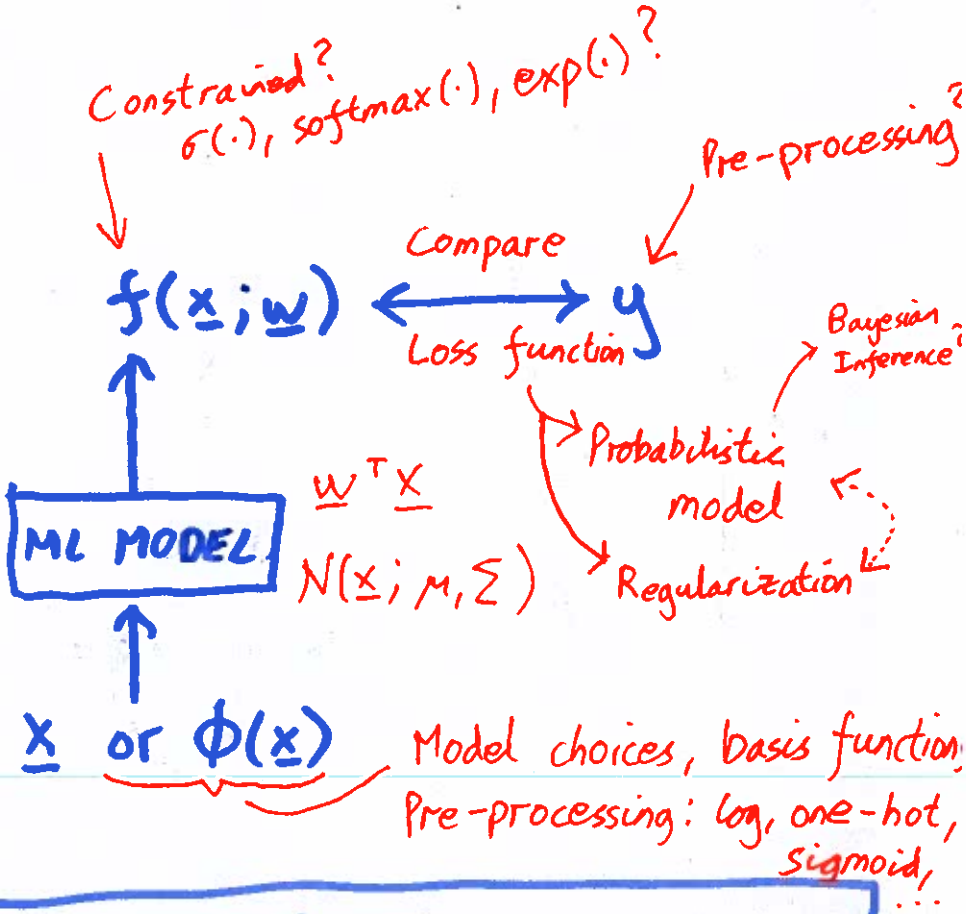


Training, validation, testing

ASSIGNMENTS

Assign 1 feedback released (check email)

Assign 2 partners assigned/confirmed today

Fill out form on assignment page by

12 noon today

if you want a partner

ML-Base, AT 5.04, 5-7pm Mon-Fr

Not usually crowded (week 1 not representative!)

Some of you found useful and welcoming

Some of you haven't :/ Not their job to give c/w hints, sorry

Useful: If you're prepared to explain + discuss issue (then if tutor still can't help, I will!)

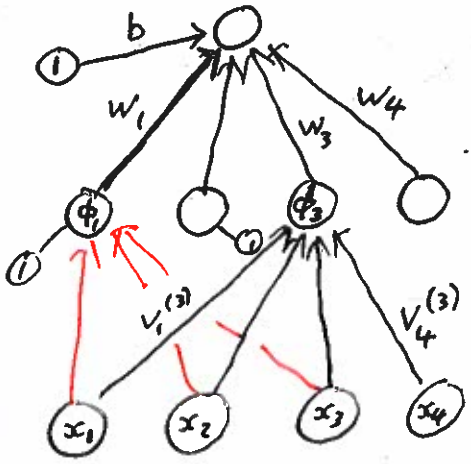
Welcoming: If you've made a nice group, please look out for those who haven't

Neural Network

First example

$$f = \sigma(\underline{w}^T \underline{\phi} + b)$$

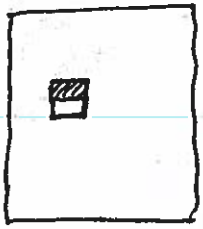
Logistic regression



$\underline{\phi}$

$$\phi_k = \sigma(\underline{v}^{(k)T} \underline{x} + b^{(k)})$$

$\underline{x}$



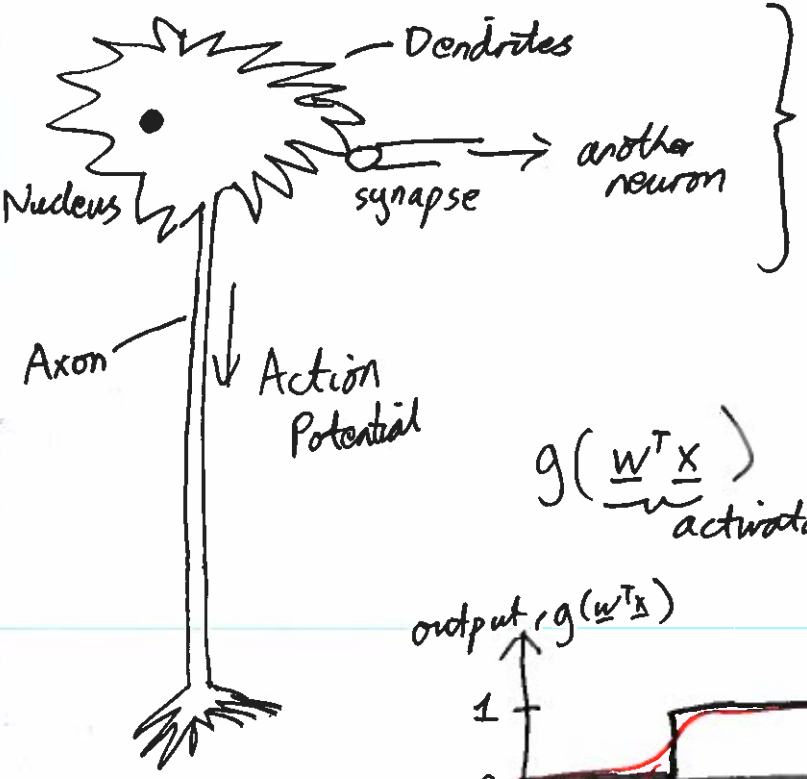
Fit $\{\underline{w}, b, \{\underline{v}^{(k)}, b^{(k)}\}\}$

with a gradient-based optimizer.

Match f to training y's using some loss.

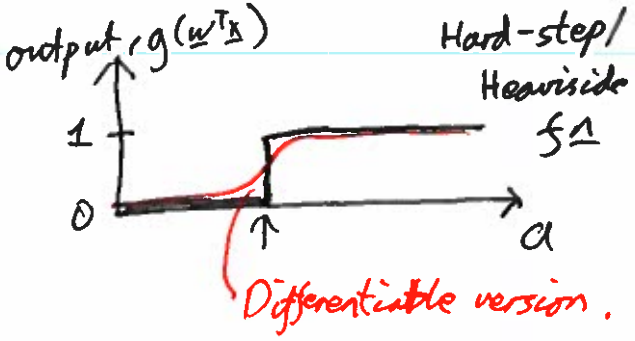
Why "Neural Network"? (non-examinable)

Neuron = Nerve cell



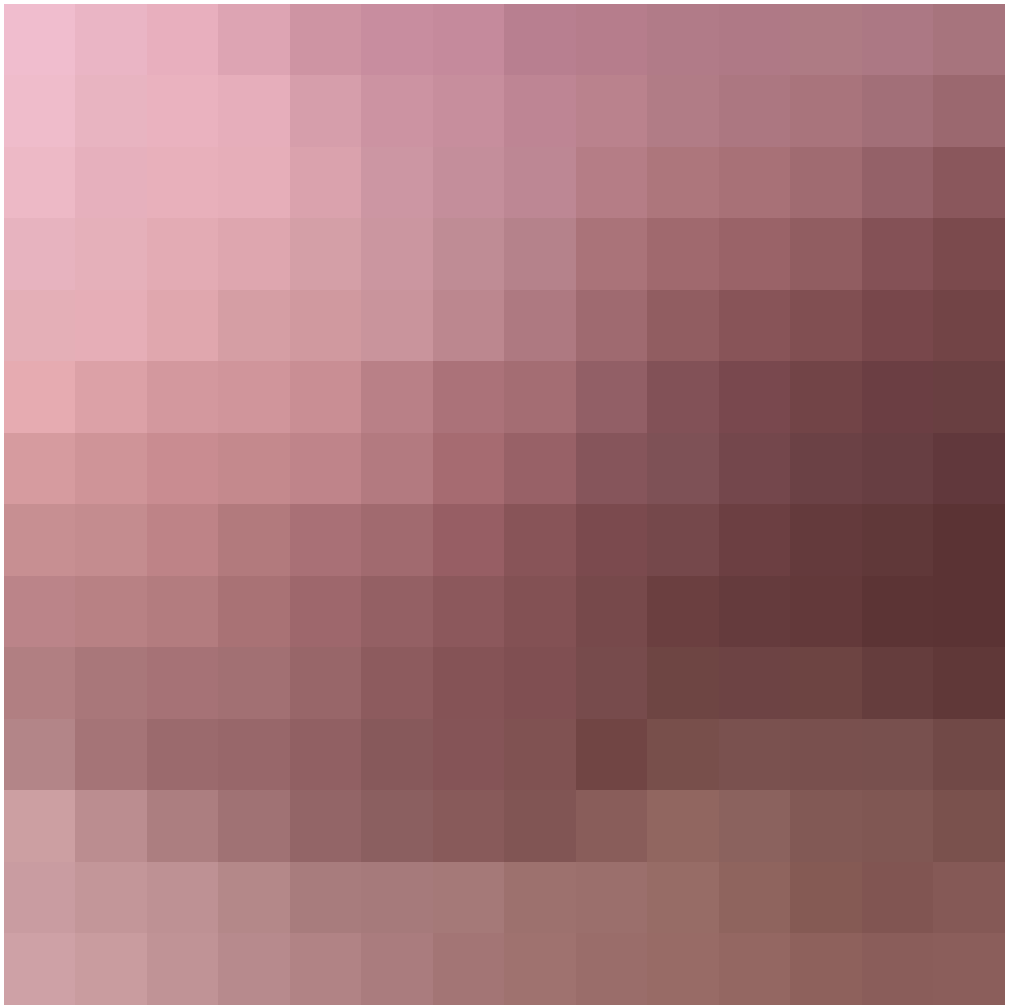
$g(\underline{w}^T\underline{x})$

activation, a



Vision in the brain

14×14 **patch of retinal pixels:** **(cartoon)**



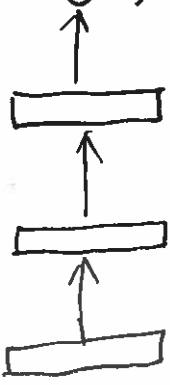
How many V1 neurons connect to each patch?

Vision in the brain

$\sim 100,000$ V1 neurons connect to 14×14 pixels in macaques

via Bruno Olshausen `http://redwood.berkeley.edu/bruno/papers/CNS2010-chapter.pdf`

Feed-forward Neural Networks



$$\underline{f} = g^{(3)}(W^{(3)}\underline{h}^{(2)} + b^{(3)})$$

$$\underline{h}^{(2)} = g^{(2)}(W^{(2)}\underline{h}^{(1)} + \underline{b}^{(2)})$$

$$\underline{h}^{(1)} = g^{(1)}(W^{(1)}\underline{x} + \underline{b}^{(1)})$$

If f is scalar $W^{(3)}\underline{h}^{(2)} = \underline{w}^{(3)T}\underline{h}^{(2)}$

Homework: try a neural function
use random weights
and plot $f(x)$

Other architectures possible

"Skip connections"

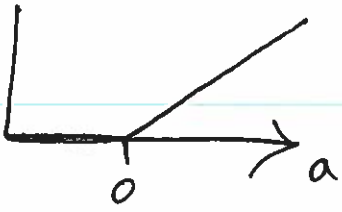
Special case: $\underline{h}^{(l)} = g^{(l)}(W^{(l)}\underline{x} + b^{(l)}) + \underline{x}$

"residual layer"

Non-linearities, g

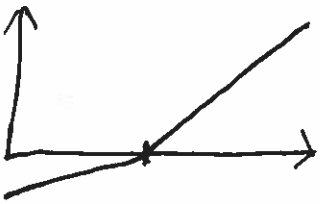
Sigmoidal function σ

ReLU: Rectified Linear Unit



$\text{Relu}(a) = \max(a, 0)$

PReLU



$$f(a) = \begin{cases} a & a > 0 \\ sa & a \leq 0 \end{cases}$$

s ↑ Parameter

Categorical data "words"

Could do:

$\underleftarrow{D}\rightarrow$ one-hot

$\underline{x} = [0\ 0\ 0 \ldots 1 \ldots 0\ 0\ 0]$

$\underline{h}^{(1)}$ can only take on D values.

$\underline{h} = \text{embedding}(x, W) = W_{:,x}$

$x \in \{1, 2, \ldots D\}$