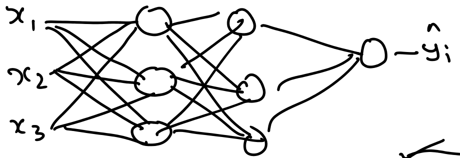
Agenda - Theory part + coding < Regression / Classification

Biological Neuron VS Artificial Neuron



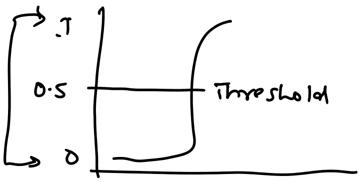
→ perceptron model - 1957

Logistic Regression - GLM

- Maximum Likelihood Estimation
- Log-loss Function
- binary class
- Multiclass
  - "OVR" → Sigmoid
  - Multinomial → softmax

$$p = \frac{1}{1+e^{-z}}$$

$$\sigma(z) = \frac{e^{z}}{\sum_{i} e^{z_i}}$$



< -∞ / +∞

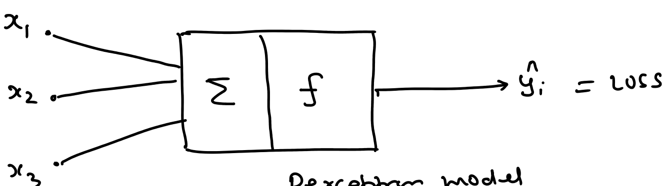
> 100 → 1
< -100 → 0

Perceptron - only one hidden layer

x_1	x_2	x_3	y per	Logistic
1	1	1	0	1
2	2	2	1	1
3	3	3	1	1
0	0	0	0	0

case 1 - Logistic Reg
↓ 0.5, more!
1

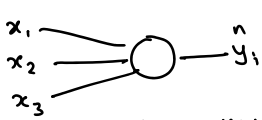
Perceptron
$x_1 + x_2 + x_3 > 5$
y = 1 otherwise 0.



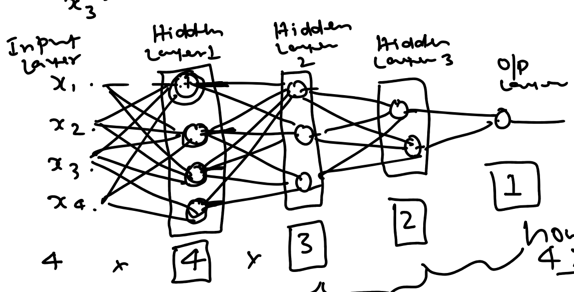
Perceptron model

n-hidden = MultiLayer Perceptron - ANN
DNN

Multilayer perceptron (MLP) < Input Layer / hidden Layer / output



= Single Layer



MLP

how many bias
4 + 3 + 2 + 1
= 10

Total per = W + B
= 36 + 10
= 46

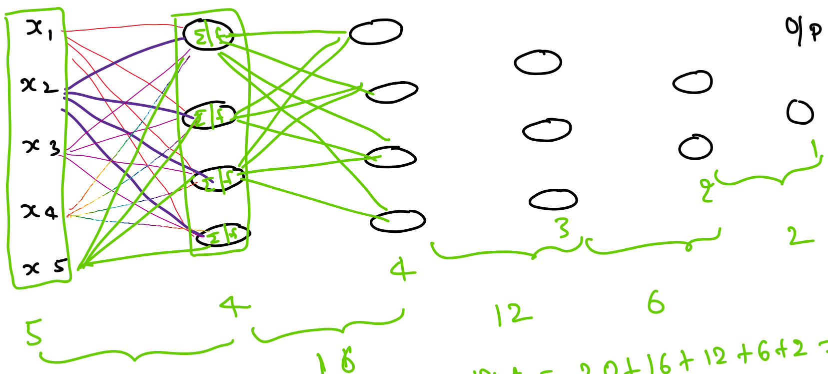
how much total weight
4 × 4 + 4 × 3 + 3 × 2 + 2 × 1
= 16 + 12 + 6 + 2 = 36

$4 \times \boxed{4} \times \boxed{3} \quad \boxed{2} \quad \boxed{1}$

$\underbrace{16} + \underbrace{12} + \underbrace{6} + 2 = 36$

→ Keras tuner
Hyperparam tuner

I/P HL1 HL2 HL3 HL4



Total Weight = 20+16+12+6+2 = 56

Total Bias = 4+4+3+2+1 = 14

Total Param = Total Weight + Total Bias

= 56 + 14 = 70 B

$\boxed{y = mx + c}$ Given, slope, Intercept

$\boxed{\hat{y} = \sum_{i=1}^{n} (w_i^T x_i + B)}$ weight, Bias

Given dataset - Titanic

Age, Gend, Fam, Per

DV = Survived

$\sum_{i=1}^{n} (y - \hat{y}_i)^2$ $y - (w_i^T x$

Q: Why should we care about MLP?

Ans :- (a) Biological Inspiration → Neuron Structure Network

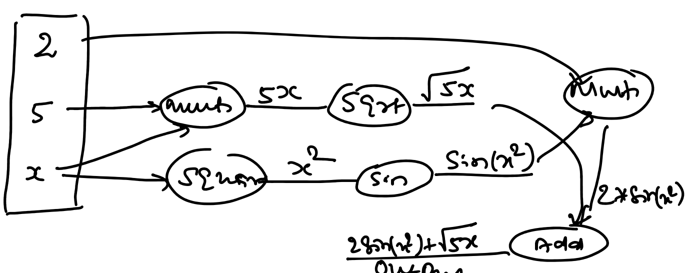
Neuroscience → millions/billions of neuron connection - human brain

(b) Mathematical Argument :-

Regression :- $\boxed{y = f(x)}$ $\boxed{2 * \sin(x^2) + \sqrt{5x}}$

f_1 = add()
f_2 = square()
f_3 = sqrt()
f_4 = Sin()
f_5 = multi()

Input non convex



NOTE :- By using MLP, we can come-up with complex math function. MLP- Structure

high-school :- Functional composition

$\boxed{f(g(x)) = g(f(x))}$ $\boxed{f \circ g(x) = g \circ f(x)}$

$\boxed{F(x) = 2 * \sin(x^2) + \sqrt{5x}}$

$f1\left(f5\left(2, f4(f2(x))\right), f3\left(f5(5,x)\right)\right)$

$5 * x = f5(5, x)$

$\sqrt{5x} = f3(f5(5,x))$

$x^2 = f2(x)$

$\sin(x^2) = f4(f2(x))$

NOTE :- MLP :- Graphical way of

$x^2 = f2(x)$

$\sin(x^2) = f4(f2(x))$

$2 * \sin(x^2) = f5(2, f4(f2(x)))$

NOTE :- MLP :- Graphical way of representation :- (fog, gof)

Functional composition

Powerful, elegant, useful — DL

Caveat ←

overfitted

---

* NOTATION

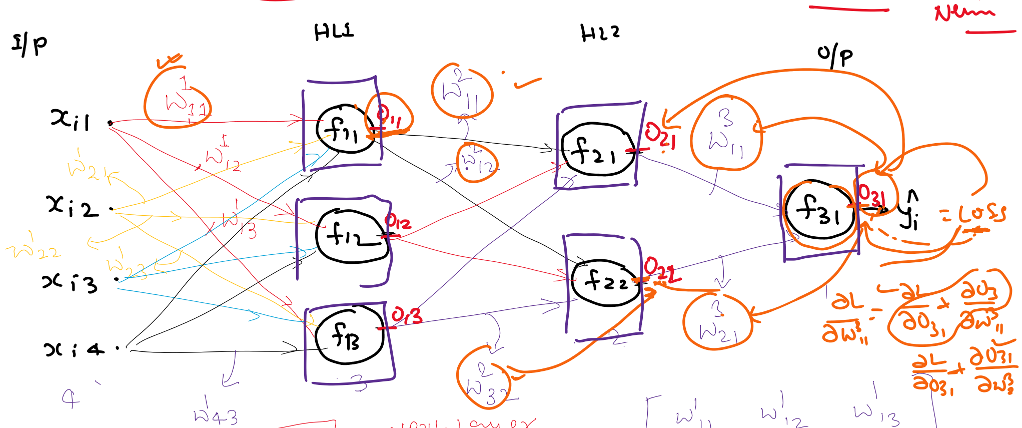
$D = \{x_i, y_i\}$, $x_i \in \mathbb{R}^4$, $y_i \in \mathbb{R}$ — Regression

x_{ij} → Feature, point

O_{ij} → index, output, Layer

f_{ij} → index, Feature — Bias, Layer

Bias = Total Neuron



Weight :- w^k_{ij} → Next Layer, → To, From

$$W_1 = \begin{bmatrix} w^1_{11} & w^1_{12} & w^1_{13} \\ w^1_{21} & w^1_{22} & w^1_{23} \\ w^1_{31} & w^1_{32} & w^1_{33} \\ w^1_{41} & w^1_{42} & w^1_{43} \end{bmatrix}_{4\times3} \quad 12$$

$$W_2 = \begin{bmatrix} w^2_{11} & w^2_{12} \\ w^2_{21} & w^2_{22} \\ w^2_{31} & w^2_{32} \end{bmatrix}_{3\times2} \quad 6$$

$$W_3 = \begin{bmatrix} w^3_{11} \\ w^3_{21} \end{bmatrix}_{2\times1} \quad 2$$

Total weight : 20

Fully-connected Layer → Dense Layer

V.V. Significant — knowledge purpose; Interview — Backpropagation

Training on Multilayer Perceptron : Chain Rule

Backpropagation Algorithm

Rule 1 :- Define Loss function (exam-Reg)

Memoizat

→ it is not

it is cal

The core

is this —

→ If ther

~~that~~ is

repeat

→ it's a

if once

for rev

→ This co

tion

memorization.

led memoization

Dynamic programming

idea of memoization

e is any operation
used many times
cally

good idea to compute
and then store it
ing purpose
cept is called

Rule 1 :- Define Loss Function (exam-Reg)

$$\boxed{L = \underbrace{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}_{\text{SQ-loss fn}} + \text{regularization}}$$

Rule 2 :- Optimization - Reduce the Loss

arg min
$w_0, w_1, w_2, w_3 \ldots$

Rule 3 :- Gradient descent method

- Batch GD / vanilla GD
- SGD
- MBGD
- → Momentum
- → NAG
- → Adagrad
- → Adadelta
- ** → Adam - 2014 ✓
- → RMSprop
- → Adamax
- → nadam etc.

(a) initialised weight
↳ lots of method
↳ xavier/Glorot { Normal, uniform }
↳ He-init

(b) update the weight

$$\boxed{\left(w_{ij}^{k}\right)_{new} = \left(w_{ij}^{k}\right)_{old} - \eta * \frac{\partial L}{\partial w_{old}}}$$

η → eta / learning rate < 0, 1

$\frac{\partial L}{\partial w_{old}}$ → V.V. Imp → Back chain Rule

(c) Perform updates till the convergence

NOTE :- Convergence means the new weight & old weight are very close to each other and no changes ~~happen~~ after doing multiple iteration

→ epochs - is nothing but iteration from Backprop to forward propagation. epoch is just a fancy word

$$\boxed{1 \text{ epoch} = 1 \text{ Backward Prop} + 1 \text{ Forward prop}}$$

Calculate $\boxed{\frac{\partial L}{\partial w_{old}}}$

$$\boxed{\frac{\partial y}{\partial x} = \frac{\partial y}{\cancel{\partial u}} \cdot \frac{\cancel{\partial u}}{\cancel{\partial v}} \cdot \frac{\cancel{\partial v}}{\partial x}}$$

$$\frac{\partial L}{\partial w_{11}^{3}} = \boxed{\frac{\partial L}{\partial O_{31}}} * \frac{\partial O_{31}}{\partial w_{11}^{3}}$$

$$\frac{\partial L}{\;} = \boxed{\frac{\partial L}{\;}} * \frac{\partial O_{31}}{\partial w^3}$$

} Chain Rule
(in calculus

→ init coo...
" Mem...

Back...

objective To ...

we...

how :- ...

Wnew = ...

...cept is called "...moization"

...propogation

...update weight & Bias so that we can set less (zero) loss value.

Chain Rule + Memoization

$$W_{new} = W_{old} - \eta \times \frac{\partial L}{\partial W_{old}}$$

$$\frac{\partial L}{\partial w^3_{11}} = \boxed{\frac{\partial L}{\partial O_{31}}} * \frac{\partial O_{31}}{\partial w^3_{11}}$$

Chain Rule

$$\frac{\partial L}{\partial w^3_{21}} = \boxed{\frac{\partial L}{\partial O_{31}}} * \frac{\partial O_{31}}{\partial w^3_{21}}$$

in calculus

$$x \rightarrow f(x) \rightarrow h, \quad x \rightarrow g(x) \rightarrow h$$

$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial f} * \frac{\partial f}{\partial x} + \frac{\partial L}{\partial g} * \frac{\partial g}{\partial x}$$

$$\frac{\partial L}{\partial w^2_{11}} = \boxed{\frac{\partial L}{\partial O_{31}}} * \frac{\partial O_{31}}{\partial O_{21}} * \frac{\partial O_{21}}{\partial w^2_{11}}$$

$$\frac{\partial L}{\partial w^2_{22}} = \boxed{\frac{\partial L}{\partial O_{31}}} * \frac{\partial O_{31}}{\partial O_{22}} * \frac{\partial O_{22}}{\partial w^2_{22}}$$

$$\frac{\partial L}{\partial w^1_{11}} = \boxed{\frac{\partial L}{\partial O_{31}}} * \frac{\partial O_{31}}{\partial O_{21}} * \frac{\partial O_{21}}{\partial O_{11}} * \frac{\partial O_{11}}{\partial w^1_{11}} + \frac{\partial L}{\partial O_{31}} * \frac{\partial O_{31}}{\partial O_{22}} * \frac{\partial O_{22}}{\partial O_{11}} * \frac{\partial O_{11}}{\partial w^1_{11}}$$

$$= \boxed{\frac{\partial L}{\partial O_{31}}} * \boxed{\frac{\partial O_{11}}{\partial w^1_{11}}} \left\{ \frac{\partial O_{31}}{\partial O_{21}} * \frac{\partial O_{21}}{\partial O_{11}} + \frac{\partial O_{31}}{\partial O_{22}} * \frac{\partial O_{22}}{\partial O_{11}} \right\}$$

Chain Rule

$$\frac{\partial L}{\partial B_{11}}$$

$$= \frac{\partial L}{\partial O_{31}} * \frac{\partial O_{31}}{\partial O_{21}} * \frac{\partial O_{21}}{\partial O_{11}} * \frac{\partial O_{11}}{\partial B_{11}} +$$

$$\frac{\partial L}{\partial O_{31}} * \frac{\partial O_{31}}{\partial O_{22}} * \frac{\partial O_{22}}{\partial O_{11}} * \frac{\partial O_{11}}{\partial B_{11}}$$