" Lecture 23: Basics of deep learning "

- Revision of previous lecture --
- why we move from traditional ML → deep learning

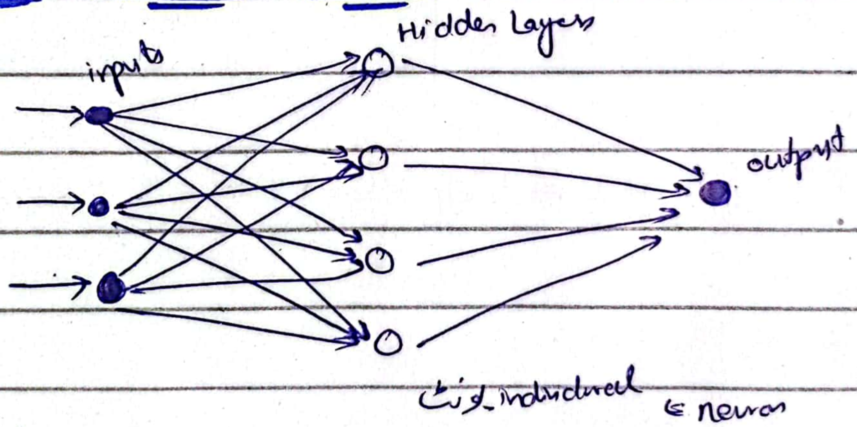
Limitation of their performance ← → feature extraction
→ high dimensional data
→ unstructured data

- Reasons for popularity
  ↳ data ↳ compute power ↳ algorithm

o difference

o short coming

Simple Neural Network



(i) input layer

neuron receive input (features)

(ii) Hidden layer

computes and pass the output to next layer

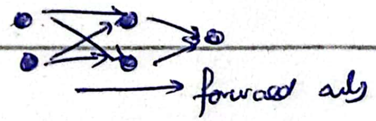
(iii) output

The final layer

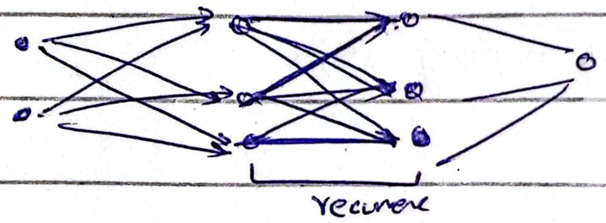
Types of Neural Network

Roughly
① Shallow Neural Network (usually 1 hidden layer)
② Deep Neural Network (multiple hidden layer)

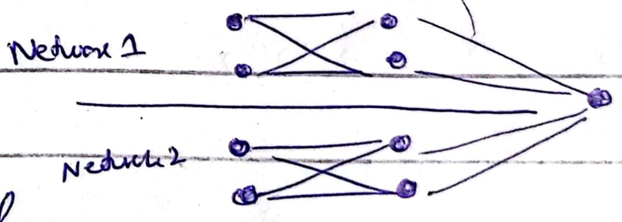
① **feed forward** (that only allow their nodes to pass information to forward nodes)



② **Recurrent** (can go backwards, allowing output from some nodes to impact input of preceding nodes)



③ **Modular Network** (combine two/more neural network in order to arrive output

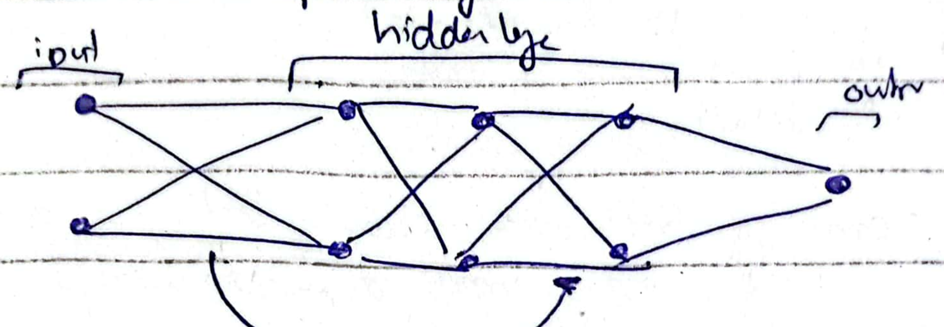


Basic Terminologies

- Training of model

④ Residual Neral Network

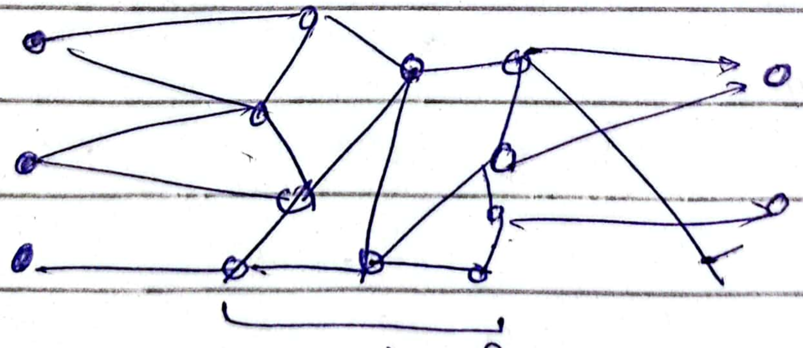
(allow data to skip ahead via a process called identity mapp, combining the output from early layers with the output layer.



identity mapping skips ahead

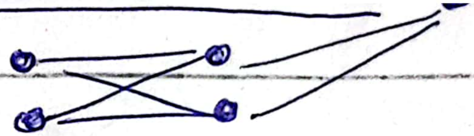
⑤ Liquid State Machine

(Neural networks features node that are randomly connected to each other)



reservoir of randomly nodes

Basic Terminologies



- Training of model

  (when data is given a model try to learn the pattern)

⇒ Hyper-parameters	Parameters
(which we have to set before the traning so model can be train)	⇒ (which model learn during the training)
⇒ Not a part of trained model & hence value not saved.	⇒ The estimated values is saved with the trained model.
⇒ External to model	⇒ part of model
Example	Example
① Learning rate ③ Batch size ② number of layers	① Co-efficients in linear regression ① weights in neural network

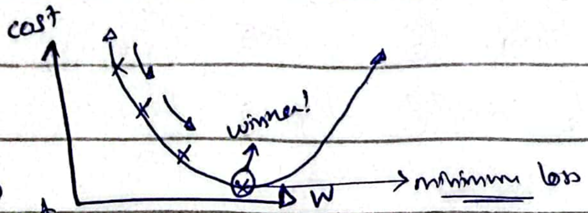
loss ⇒ penalty
→ The more mistake more loss

Hyperameters

(1) learning rate :

(a hyper-parameter that controls the step size during gradient descent)

determines how quickly model adjusts its parameters in directon that reduce the loss.

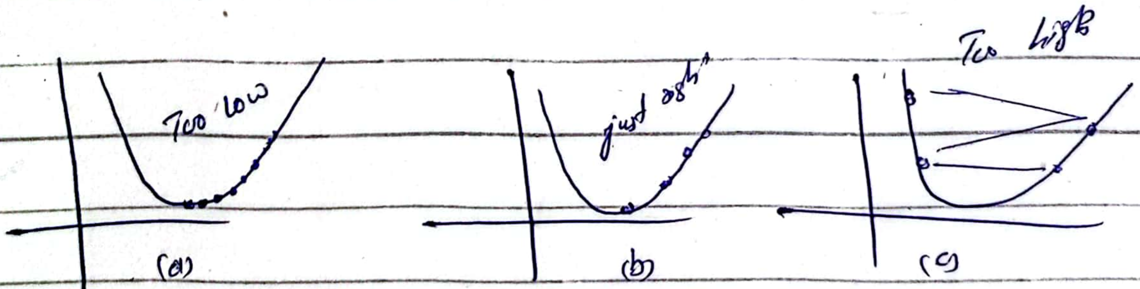


|| goal lowest point ||

Gradient descent ↳ helps us to go in right directon

A higher learning rate might led to faster convergnce but risks over-shooting.

⇒ A slower rate slow dow converge



small learning rate require many updates before reaching minimum point

swiftly reaches the minimum point

cause drastic updates which leed to divergent behav

② Batch Size

Batch ⇒ a sub-set of the training dataset used used in each iteraton of traning process.

⇒ we divide the entire dataset into smaller batches

factors
(hardware capacity, computational power, complexity of algo)

Total data ⇒ 1000

Batch size = 100 → iterator ⇒ 10

✓ ہم اتنی ڈیٹالیف ماڈل کو ایک ساتھ دے سکتے ہیں

③ Number of epoch

(one complete round of data)

⇒ The number of times the entire training dataset is seen by the model during training.

Too few epochs ⇒ underfitting

Too many epochs → overfitting

⇒ Tensorflow Playgrond