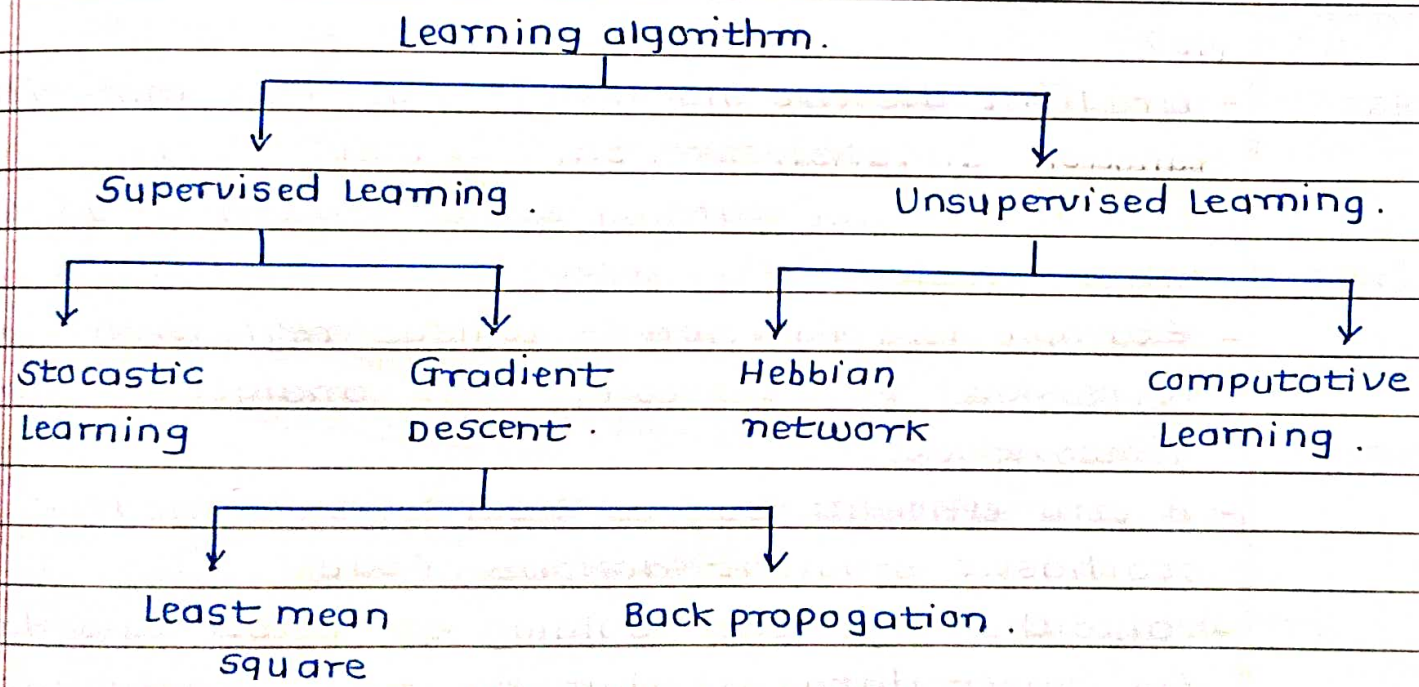
30/01/2024

classmate
Date
Page

Assignment No :- 02

Que1 Explain learning algorithm with respect to artificial neural network ?

Ans:
- Learning is a process where unknown ANN parameters are adapted through continuous process of stimulation from the environment.
- Learning is determined by the way how the change of parameters takes place.
- A set of rules that are solution to the learning problem is called a learning algorithm.
- Error correction belongs to the supervised learning paradigm.



i) Supervised Learning :-

- supervised learning algorithm desired output is given.
- It is not possible to learn larger and more complex models than with supervised learning.
- Use training data to infer model.

- Every input pattern that is used to train the network is associated with an output pattern.
- Trying to predict a function from labeled data.

ii) Unsupervised Learning :-

- In unsupervised learning algorithm desired output is not given.
- It is possible to learn larger and more complex model with unsupervised learning.
- No training data is used.
- The target output is not presented to the network
- Try to detect interesting relations in data.

Ques) Explain gradient descent algorithm with suitable example ?

Ans.
- Gradient descent algorithm is used to minimized the function.
- Much of machine learning can be written as an optimization problem.
- Example loss function :- Logistic regression, linear regression, principle component analysis, neural network loss.
- A very efficient way to train logistic model is with stochastic gradient descent (SGD).
- One challenge with training on power law data (i.e most data) is that the terms in the gradient can have very difficult strengths.
- The idea behind stochastic gradient descent is iterating a weight updates based on the gradient of loss function :

$$\overline{w}(k+1) = \overline{w}(k) - \gamma \nabla L(\overline{w})$$

- Logistic regression is designed as a binary classifer (output say {0,1}) but actually outputs the probability that the input instance is in the "1" class.
- A logistic classifier has the form :

$$P(X) = \frac{1}{1 + \exp(-XB)}$$

Where,

$X = (X_1, X_2, X_3, \ldots, X_n)$ is a vector of features.

- The process of finding most optimal hyperparameters in machine learning is called hyperparameter optimisation.

Ques) Explain Hebbian Learning / Describe hebbs algorithm.

Ans. - In 1949, Donald Hebb proposed one of the key ideas in biological learning, commonly known as Hebb's law.
- Hebb's law states that if neuron i is near enough to excite neuron j and repeatedly participates in its activation, the synaptic connection between these two neurons is strengthened and neuron j becomes more sensitive to stimuli from neuron i.
- Hebb's law provides the basis for learning without a teacher. Learning here is a local phenomenon occurring without feedback from the environment.

$$\Delta W_{ij}(P) = F[Y_i(P), X_i(P)].$$

- As a special case, we can represent Hebb's law as follows :-

$$\Delta W_{ij}(P) = \alpha\, y_i(P)\, X_i(P).$$

Where α is the learning rate parameter. This equation is referred to as the activity product rule.

$$\Delta W_{ij}(P) = \alpha\, Y_i(P)\, X_i(P) - \varphi\, Y_i(P)\, W_{ij}(P)$$

$\therefore$ φ is forgetting factor.

- Hebbian learning algorithm :-

- step 1 :- Initialisation.

set initial synaptic weights and threshold to small random values, say in an interval [0,1].

- step 2 :- Activation.

Compute the neuron output at iteration p.

$$y_i(p) = \sum_{i=1}^{n} x_i(p)\, w_{ij}(p) - \theta_j$$

where,

n = No. of neuron inputs.

- step 3 :- Learning.

Update the weights in the network.

$$w_{ij}(p+1) = w_{ij}(p) + \Delta w_{ij}(p)$$

∴ w = weight.

$$w_{ij}(p) = \alpha\, y_i(p)\, [\Delta x_i(p) - w_{ij}(p)]$$

- step 4 :- Iteration.

Increase iteration p by one, go back to step 2.

Que4) Explain gradient descent algorithm using sigmoid function

Ans. - A sigmoid function is a mathematical function with a characteristic "S"-shaped curve or sigmoid curve. It transforms any value in the domain $(-\infty, \infty)$ to a number between 0 and 1.

- Although the sigmoid function is prevalent in the context of gradient descent, the gradient of the sigmoid function is in some cases problematic. The gradient vanishes to zero for very low and very high input values, making it hard for some models to improve.

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

Where,

x = Input to the sigmoid function.

e = Euler's number

- The derivative of the sigmoid function is :-

$$\left[\frac{d}{dx}\sigma(x) = \sigma(x)\,(1-\sigma(x))\right]$$

$$\frac{d}{dx}\sigma(x) = \frac{d}{dx}\left[\frac{1}{1+e^{-x}}\right]$$

$$= \frac{d}{dx}\left[(1+e^{-x})^{-1}\right]$$

$$= \frac{d}{dx}\left[u^{-1}\right], \quad u = 1+e^{-x}.$$

$$= \frac{d}{dx} -u^{-2}\frac{du}{dx}$$

$$= -(1+e^{-x})^{-2} \cdot \frac{d}{dx}\left[1+e^{-x}\right]$$

$$= -(1+e^{-x})^{-2} \cdot -e^{-x}$$

$$= \frac{-e^{-x}}{(1+e^{-x})^2}$$

$$= \frac{1}{(1+e^{-x})} \cdot \frac{-e^{-x}}{1+e^{-x}}$$

$$= \frac{1}{(1+e^{-x})} \cdot \frac{(1+e^{-x})-1}{1+e^{-x}}.$$

$$= \frac{1}{(1+e^{-x})} \cdot \left(\frac{1+e^{-x}}{1+e^{-x}} - \frac{1}{1+e^{-x}}\right)$$

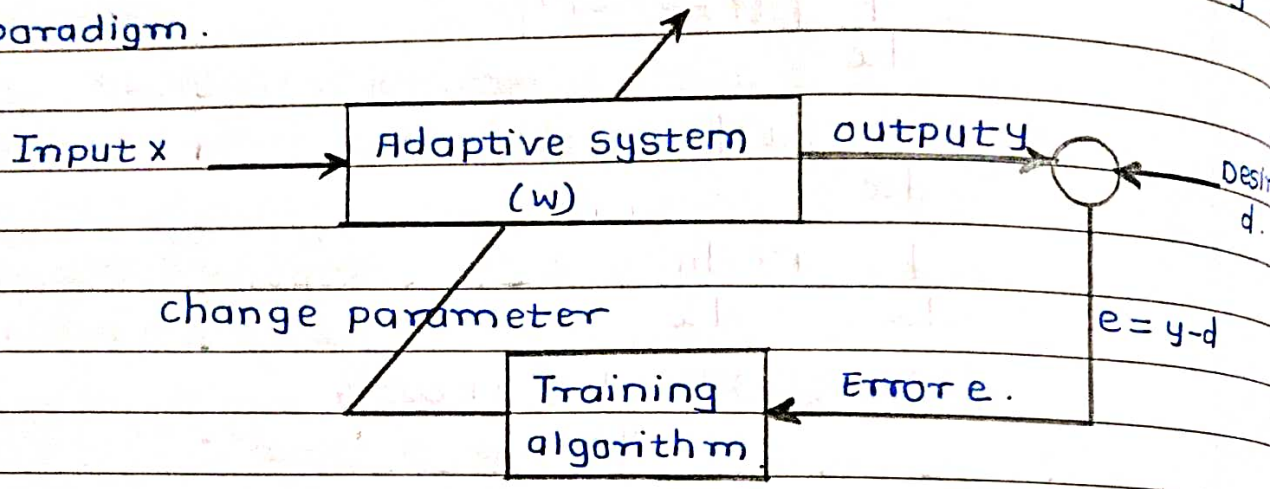
$$= \frac{1}{(1+e^{-x})} \cdot \left(1 - \frac{1}{1+e^{-x}}\right)$$

$$= \sigma(x) \cdot (1-\sigma(x)).$$

$$\therefore \frac{d}{dx}\sigma(x) = \sigma(x) \cdot (1-\sigma(x)).$$

Que 5) Explain Error correction in gradient descent rule.

Ans:
- Learning is a process where unknown ANN parame
are adapted through continuous process of stimulati
from the environment.
- Learning is determined by the way how the change
parameters takes place. A set of rules that are solu
to the learning problem is called a learning algorithm.
- Error correction belongs to the supervised learning
paradigm.



- Error correction learning -

- Error values can be used to update or adjust using
algorithm like Feed back or back propogation.

$\therefore e = T - Y.$

Where,

e = error.

T = Threshold.

Y = output.

- The goal of error-correction learning is to mini
an error function derived from error $e_k(n)$ so tha
the obtained output of all neurons approximate
the desired output is some statistical sense.
- A frequentely used error function is mean sq
error :-

$$J = E\left[\frac{1}{2}\sum_{k} e_k^2(n)\right].$$

Where,
E[·] is the statistical expectation operator and summation is for all neurons in the output layer.

Ques) Explain difference between perceptron model and adaline model.

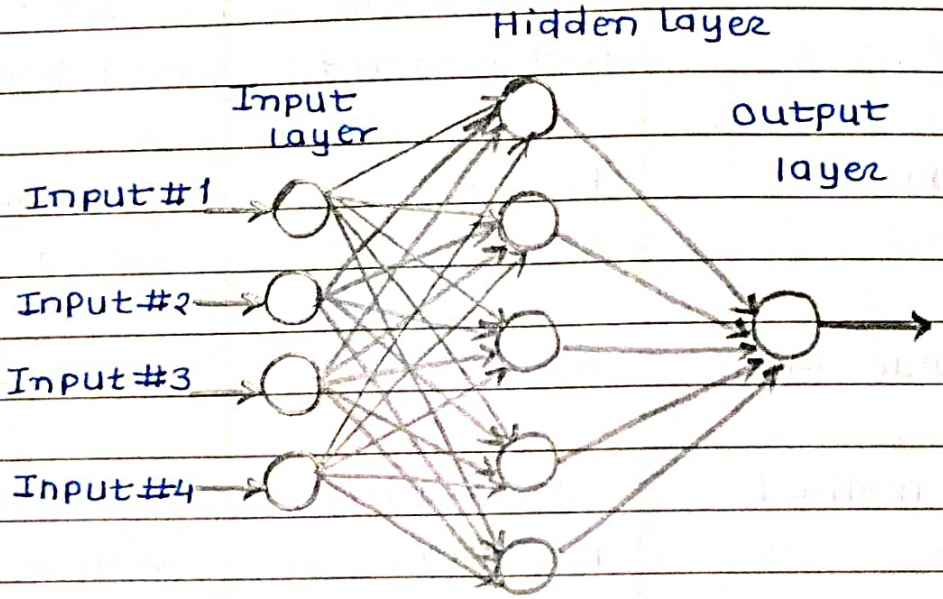
Ans:

Model	perceptron	Adaline.
i) Type	Feed Forward.	Feed Forward.
ii) Neuron layers	1 input layer 1 output layer	Many input layer 1 output layer.
iii) Input value types.	Binary input.	Binary input.
iv) learning method.	Hard limiter learning.	supervised learning.
v) Learning algorithm.	Hebb learning rule.	Gradient descent algorithm.
vi) Application.	i) Simple logical operations. ii) pattern classification.	i) Simple logical operations. ii) Regressions.

Que 1) Explain multilayer feedback neural network.

Ans. - A multilayer feedback neural network, also known as a multilayer perceptron (MLP), is a type of artificial neural network with multiple layers of interconn nodes or neurons.

- The network typically consists of an input layer, one or more hidden layers, and an output layer.
- Each connection between nodes has a associated weight, and each node applies an activation function to the weighted sum of its inputs.



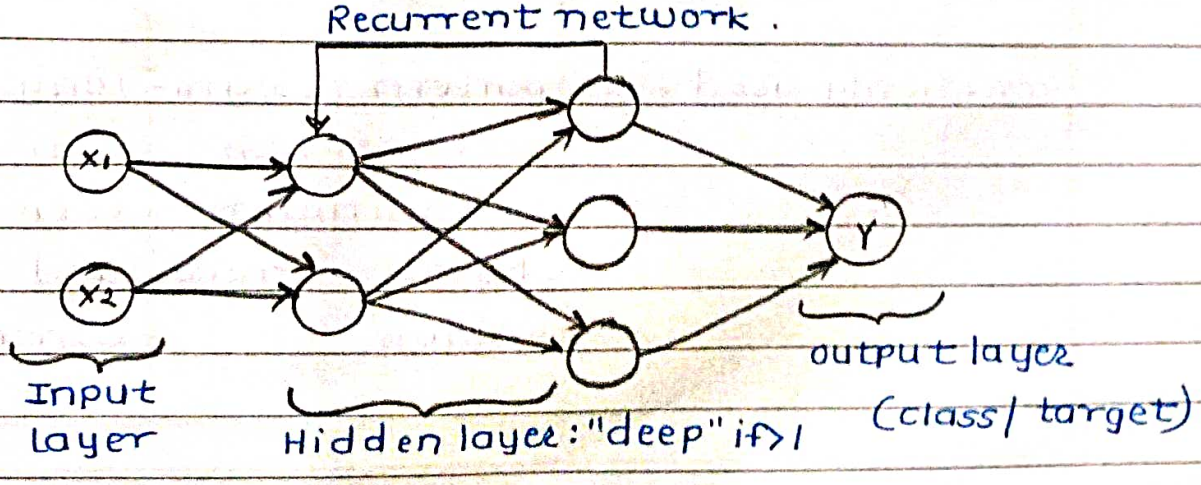
Multilayer perceptron model.

- The term "feedback" refers to the training process where the network learns by adjusting the weights based on the error between its predicted output and desired output.
- This is often done using backpropogation, where the error is propagated backward through the network, and the weights are updated accordingly to minimize the difference between predicted and actual outputs.

- The hidden layers in a multilayer feedback neural network allow for complex representations & features learning, enabling the network to capture intricate patterns in the input data.
- This architecture is widely used in various applications, including image recognition, natural language processing, and many other machine learning tasks.

Que8) Explain recurrent neural network.

Ans.
- A recurrent neural network is a type of neural network that contains loops, allowing information to be stored within the network.
- A RNN is particularly useful when a sequences of data is being processed to make a classification decision or regression estimate but it can also be used on non-sequential data.
- Recurrent neural networks are typically used to solve tasks related to time series data.
- Applications of recurrent neural network include natural language processing, speech recognition, machine translation, character-level language modeling, image classification, image captioning, stock prediction, and financial engineering.



- Recurrent neural networks can be thought of as a series of networks linked together. They often have a chain-like architecture, making them applicable for tasks such as speech recognition, language translation.
- An RNN can be designed to operate across sequence of vectors in the input, output, or both.

Ques) Difference between feed forward and recurrent neural network ?

Ans.

Comparison Attribute.	Feed-Forward neural network.	Recurrent neural network.
Single flow direction	Forward only	Bidirectional
Delay introduced	No	Yes
complexity	Low	High
Neuron independence in the same layer.	Yes	No
Speed	High	slow
commonly used for	pattern recognition, speech recognition & character recognition.	Language translation, speech-to-text conversion, and robotic control

X-O-X