



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING (DATA SCIENCE)

COURSE CODE: DJ19DSC501

DATE:

COURSE NAME: Machine Learning - II

CLASS: AY 2021-22

LAB EXPERIMENT NO.2

NAME: Bhuvu Ghosh

SAP ID: 60009210191

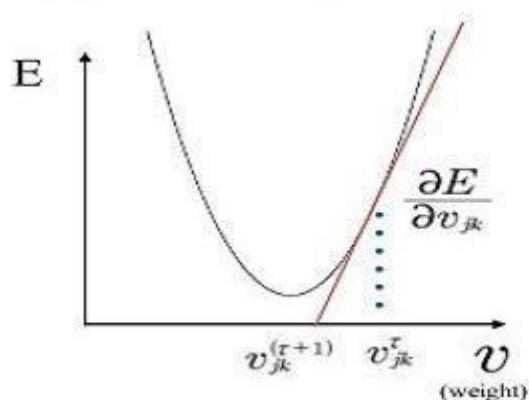
BATCH: D22

AIM / OBJECTIVE:

Implement delta learning rule using Stochastic and Batch gradient descend algorithm from scratch.

DESCRIPTION OF EXPERIMENT:

Gradient descent is a way to find a minimum in a high-dimensional space in direction of the steepest descent. The delta rule is an update rule for single layer perceptron's. It makes use of gradient descent.



$$v_{jk}^{(\tau+1)} = v_{jk}^{\tau} + \Delta v_{jk}$$

$$\Delta v_{jk} = -\eta \frac{\partial E}{\partial v_{jk}}$$

$v_{jk}^{(\tau+1)}$ new weight
 v_{jk}^{τ} current weight
 η learning rate
 E Error Function

Weight updation in delta learning

The Delta Rule uses the difference between target activation (i.e., target output values) and obtained activation to drive learning. For reasons discussed below, the use of a threshold activation function (as used in both the McCulloch-Pitts network and the perceptron) is dropped & instead a linear sum of products is used to calculate the activation of the output neuron (alternative activation functions can also be applied). Thus, the activation function is called a Linear Activation function, in which the output node's activation is simply equal to the sum of the network's respective input/weight products. The strength of network connections (i.e., the values of the weights) are adjusted to reduce the difference between target and actual output activation (i.e., error). A graphical depiction of a simple two-layer network capable of deploying the Delta Rule is given in the figure above w (Such a network is not limited to having only one output node):

During forward propagation through a network, the output (activation) of a given node is a function of its inputs. The inputs to a node, which are simply the products of the output of



**SHRI VILEPARLE KELAVANI MANDAL'S
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING**
(Autonomous College Affiliated to the University of Mumbai)
NAAC ACCREDITED with "A" GRADE (CGPA : 3.18)



preceding nodes with their associated weights, are summed and then passed through an activation function before being sent out from the node. Thus, we have the following:

This delta learning rule is also known as the Least Mean Squares (LMS) or Widrow-Hoff rule. The basic delta rule is given by

$$\Delta w_{ij} = \alpha (t_j - y_j) x_i$$

Where

t_j = the teacher
value (or desired
value) for unit j . y_j

= the actual

in other words

output for unit j .

α = learning rate.

$$\Delta w_{ij} = \alpha \delta x_i$$

Where $\delta = t_j - y_j$

i.e. the difference between the desired or target output and the actual output . The delta rule modifies the weights appropriately for both continuous and binary inputs and outputs .

The delta rule minimizes the squares of the differences between the actual and the desired O/P values.

the squared error for a particular training

$$\text{pattern } E = \sum_j (t_j - y_j)^2$$

Where E is a fn. of all the weights . The gradient of E is a vector consisting of the partial derivatives of E with respect to each of the weights. This vector gives the direction of most rapid increase in E , the opposite direction gives the direction of most rapid decrease in the error . The error can be reduced most rapidly by adjusting the weight w_{ij} in the direction of $-\partial E / \partial w_{ij}$.

Generalized delta rule:-

In a multilayer network, the determination of the error is a recursive process which starts with the O/P units and the error is back propagated to the input unit. Therefore the rule is called the error back propagation BP.

$$\Delta w_{ij} = \alpha \delta_j x_i$$

$$\delta_j = (t_j - y_j) f'(y - \text{in}_j).$$

where

$$y - \text{in}_j = \sum w_j x_i + b$$

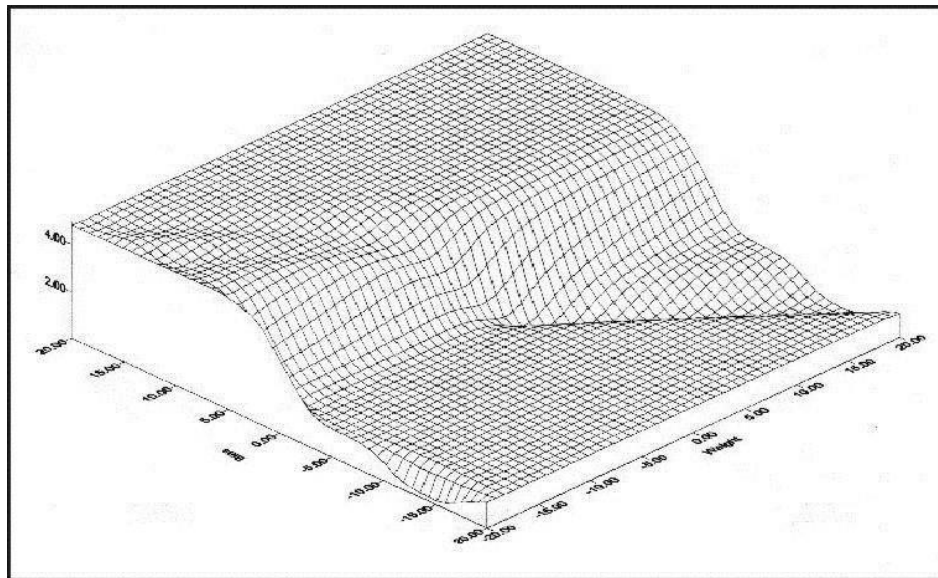
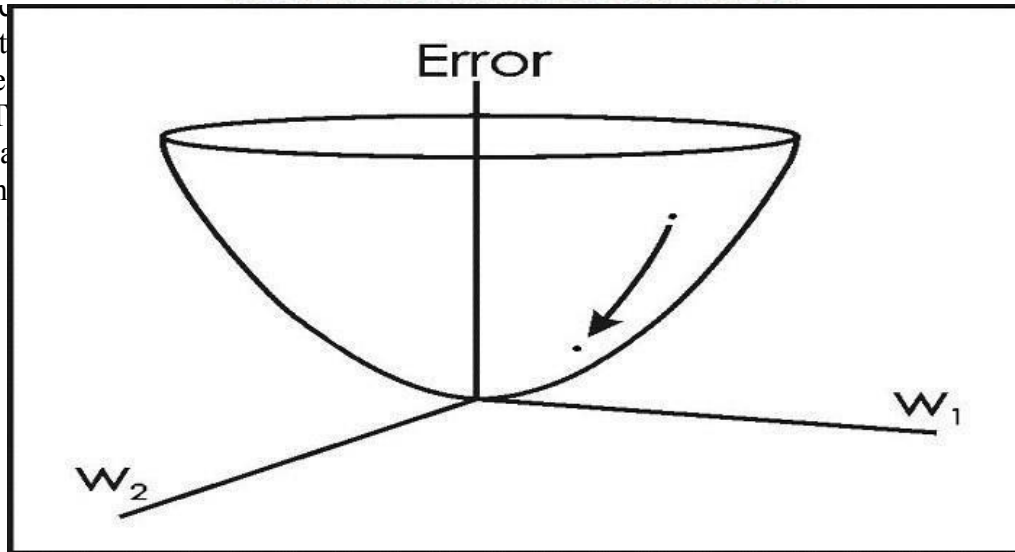


SHRI VILEPARLE KELAVANI MANDAL'S
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC ACCREDITED with "A" GRADE (CGPA : 3.18)



For i
Error func
of weight
given we
weight. T
between a
(i.e., train

ication
plied to a
t to that
ferences
g pattern





Algorithm Steps, which involves:

- **Selecting Input & Output:** The first step of the delta learning algorithm is to choose an input for the process and to set the desired output.
- **Setting Random Weights:** Once the input and output are set, random weights are allocated, as it will be needed to manipulate the input and output values. After this, the output of each neuron is calculated through the forward propagation, which goes through:
 - Input Layer
 - Output Layer
- **Error Calculation:** This is an important step that calculates the total error by determining how far and suitable the actual output is from the required output. This is done by calculating the errors at the output neuron.
- **Error Minimization:** Based on the observations made in the earlier step, here the focus is on minimizing the error rate to ensure accurate output is delivered.
- **Updating Weights & other Parameters:** If the error rate is high, then parameters (weights and biases) are changed and updated to reduce the rate of error using the delta rule or gradient descent. This is accomplished by assuming a suitable learning rate and propagating backward from the output layer to the previous layer. Acting as an example of dynamic programming, this helps avoid redundant calculations of repeated errors, neurons, and layers.
- **Modeling Prediction Readiness:** Finally, once the error is optimized, the output is tested with some testing inputs to get the desired result. This process is repeated until the error reduces to a minimum and the desired output is obtained.

Task(s) to be performed:

- a) Take the dataset with initial value of $X = [0.5, 2.5], Y = [0.2, 0.9]$
- b) Initialize a neural network with random weights.
- c) Calculate output of Neural Network:
 - i. Calculate error
 - ii. Note weight and bias changes
 - iii. Calculate loss



iv. Plot error surface using loss function verses weight, bias

- d) Perform this cycle in step c for every input output pair
- e) Perform multiple epochs of step d.
- f) Update weights accordingly using stochastic and batch gradient descend.
- g) Plot the mean squared error for each iteration 0 in stochastic and Batch Gradient Descent.
- h) Similarly plot accuracy for iterations and note the results.

SOURCE CODE:

Implement backpropagation algorithm from scratch.

```
[ ] import numpy as np
import pandas as pd
from sklearn.datasets import load_iris
from sklearn.model_selection import train_test_split
import matplotlib.pyplot as plt
```

```
[ ] data = load_iris()

X=data.data
y=data.target
```

```
[ ] y = pd.get_dummies(y).values

y[:3]

array([[1, 0, 0],
       [1, 0, 0],
       [1, 0, 0]], dtype=uint8)
```

```
[ ] x_train, X_test, y_train, y_test = train_test_split(X, y, test_size=20, random_state=4)
```

```
[ ] learning_rate = 0.1
iterations = 5000
N = y_train.size
input_size = 4
hidden_size = 2
output_size = 3
results = pd.DataFrame(columns=["mse", "accuracy"])
```

```
[ ] np.random.seed(10)

W1 = np.random.normal(scale=0.5, size=(input_size, hidden_size))

W2 = np.random.normal(scale=0.5, size=(hidden_size, output_size))
```

```
def sigmoid(x):
    return 1 / (1 + np.exp(-x))

def mean_squared_error(y_pred, y_true):
    return ((y_pred - y_true)**2).sum() / (2*y_pred.size)

def accuracy(y_pred, y_true):
    acc = y_pred.argmax(axis=1) == y_true.argmax(axis=1)
    return acc.mean()
```

```
[ ] for itr in range(iterations):

    Z1 = np.dot(x_train, W1)
    A1 = sigmoid(Z1)

    Z2 = np.dot(A1, W2)
    A2 = sigmoid(Z2)

    mse = mean_squared_error(A2, y_train)
    acc = accuracy(A2, y_train)
    results=results.append({"mse":mse, "accuracy":acc},ignore_index=True )

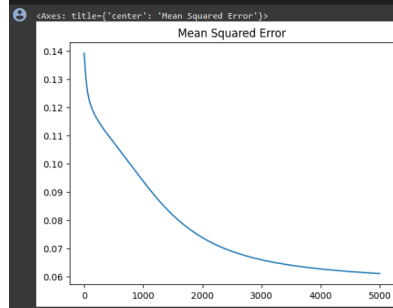
    E1 = A2 - y_train
    dW1 = E1 * A2 * (1 - A2)
```

```
E2 = np.dot(dW1, W2.T)
dW2 = E2 * A1 * (1 - A1)

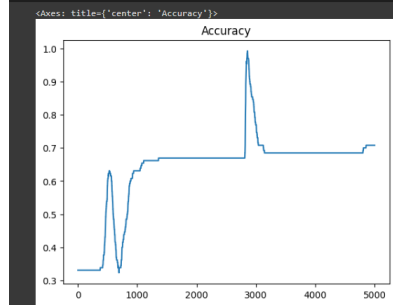
W2_update = np.dot(A1.T, dW1) / N
W1_update = np.dot(x_train.T, dW2) / N

W2 = W2 - learning_rate * W2_update
W1 = W1 - learning_rate * W1_update
```

```
results.mse.plot(title="Mean Squared Error")
```



```
results.accuracy.plot(title="Accuracy")
```



```
[ ] # Testing the model
hidden_layer_input_test = np.dot(X_test, W1)
hidden_layer_output_test = sigmoid(hidden_layer_input_test)
output_layer_input_test = np.dot(hidden_layer_output_test, W2)
output_layer_output_test = sigmoid(output_layer_input_test)

# Calculate test metrics
test_mse = mean_squared_error(output_layer_output_test, y_test)
test_acc = accuracy(output_layer_output_test, y_test)

print("Test Mean Squared Error:", test_mse)
print("Test Accuracy:", test_acc)
```

Test Mean Squared Error: 0.050542200400727086
Test Accuracy: 0.8