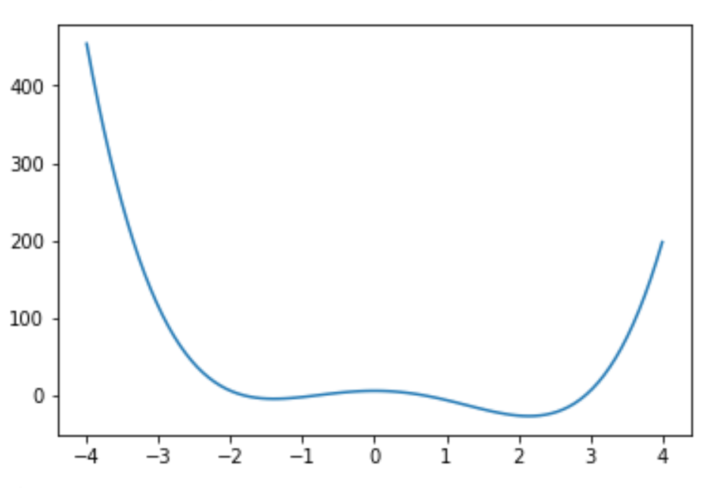
Assignment 4

Rohit Garla	Rg3365

Saturday, November 24, 2018 8:14 PM

1)
A)
Answer:
Value of x at local minimum = -1.397180
Value of x at global minimum = 2.147180



B)
When x starts with -4 and $\eta = 0.001$,

```
Before entering the iteration Value of x is: -4 Value of f is: 454
At the end of iteration 1 Value of x is: -3.488 Value of f is: 240.907412201472
At the end of iteration 2 Value of x is: -3.159231053824 Value of f is: 148.52441854620668
At the end of iteration 3 Value of x is: -2.922916422502639 Value of f is: 99.4029877988203
At the end of iteration 4 Value of x is: -2.7420316758639505 Value of f is: 70.07121494417243
At the end of iteration 5 Value of x is: -2.5977950740777596 Value of f is: 51.16573699678769
At the end of iteration 6 Value of x is: -2.479400344271616 Value of f is: 38.296442311327496
```

After running 1200 iterations,

```
Before entering the iteration Value of x is: -4 Value of f is: 454
At the end of iteration 1 Value of x is: -3.488 Value of f is: 240.907412201472
At the end of iteration 2 Value of x is: -3.159231053824 Value of f is: 148.52441854620668
At the end of iteration 3 Value of x is: -2.922916422502639 Value of f is: 99.4029877988203
At the end of iteration 4 Value of x is: -2.7420316758639505 Value of f is: 70.07121494417243
At the end of iteration 5 Value of x is: -2.5977950740777596 Value of f is: 51.16573699678769
At the end of iteration 6 Value of x is: -2.479400344271616 Value of f is: 38.296442311327496
At the end of iteration 1195 Value of x is: -1.3971808598447308 Value of f is: -4.348957724100303
At the end of iteration 1196 Value of x is: -1.3971808598447308 Value of f is: -4.348957724100303
At the end of iteration 1197 Value of x is: -1.3971808598447308 Value of f is: -4.348957724100303
At the end of iteration 1198 Value of x is: -1.3971808598447308 Value of f is: -4.348957724100303
At the end of iteration 1199 Value of x is: -1.3971808598447308 Value of f is: -4.348957724100303
At the end of iteration 1200 Value of x is: -1.3971808598447308 Value of f is: -4.348957724100303
```

The value of x has converged to the local minimum.

C)
When x starts with 4 and we do 1200 iterations

```
Before entering the iteration Value of x is: 4 Value of f is: 198
At the end of iteration 1 Value of x is: 3.68 Value of f is: 110.61233152
At the end of iteration 2 Value of x is: 3.450886144 Value of f is: 64.53629857986431
At the end of iteration 3 Value of x is: 3.276396901609702 Value of f is: 37.31076190742675
At the end of iteration 4 Value of x is: 3.138067975365072 Value of f is: 19.971643359608066
At the end of iteration 5 Value of x is: 3.0252501730040535 Value of f is: 8.322601113072949
At the end of iteration 6 Value of x is: 2.9312689375235244 Value of f is: 0.17557478693808548
At the end of iteration 1195 Value of x is: 2.1471808598447315 Value of f is: -26.611979775899698
At the end of iteration 1196 Value of x is: 2.1471808598447315 Value of f is: -26.611979775899698
At the end of iteration 1197 Value of x is: 2.1471808598447315 Value of f is: -26.611979775899698
At the end of iteration 1198 Value of x is: 2.1471808598447315 Value of f is: -26.611979775899698
At the end of iteration 1199 Value of x is: 2.1471808598447315 Value of f is: -26.611979775899698
At the end of iteration 1200 Value of x is: 2.1471808598447315 Value of f is: -26.611979775899698
```

D)
When x starts with -4 and $\eta = 0.01$

```
Before entering the iteration Value of x is: -4 Value of f is: 454
At the end of iteration 1 Value of x is: 1.12 Value of f is: -8.71561728
At the end of iteration 2 Value of x is: 1.35166976 Value of f is: -14.187225687602176
At the end of iteration 3 Value of x is: 1.5881299140655707 Value of f is: -19.554356180837104
At the end of iteration 4 Value of x is: 1.8001695002820235 Value of f is: -23.55150883046352
At the end of iteration 5 Value of x is: 1.9599549783032466 Value of f is: -25.642047221895847
At the end of iteration 6 Value of x is: 2.0585082124451546 Value of f is: -26.383081197323108
At the end of iteration 1195 Value of x is: 2.147180859844728 Value of f is: -26.61197977589969
At the end of iteration 1196 Value of x is: 2.147180859844728 Value of f is: -26.61197977589969
At the end of iteration 1197 Value of x is: 2.147180859844728 Value of f is: -26.61197977589969
At the end of iteration 1198 Value of x is: 2.147180859844728 Value of f is: -26.61197977589969
At the end of iteration 1199 Value of x is: 2.147180859844728 Value of f is: -26.61197977589969
At the end of iteration 1200 Value of x is: 2.147180859844728 Value of f is: -26.61197977589969
```

We see that x has converged to global minimum

E)
We can see that x values bounces all over the place never to converge at a single point. This is because the learning rate is too high

```
Before entering the iteration Value of x is: -4 Value of f is: 454
At the end of iteration 1 Value of x is: 47.2 Value of f is: 9689505.955200002
At the end of iteration 2 Value of x is: -82626.05440000004 Value of f is: 9.321875746621 32e+19
At the end of iteration 3 Value of x is: 451278842347294.25 Value of f is: 8.294875771953866e+58
At the end of iteration 4 Value of x is: -7.352328532672769e+43 Value of f is: 5.8442611657954316e+175
At the end of iteration 5 Value of x is: 3.1795429923047176e+131 Value of f is: nan
At the end of iteration 6 Value of x is: -inf Value of f is: nan
```

2)
A)
The pseudo code implements stochastic gradient descent (SGD)

To complete one epoch there should be 500 iterations,
To complete 100 epochs there should be $100 * 500 = 50000\ iterations$

B)
i)

$$\Delta V_{23} = -\eta \frac{d}{dV_{23}}\left[\frac{1}{2}\left(3(r_1-y_1)^2+7(r_2-y_2)^2\right)\right]$$

$$\Delta V_{23} = -\eta\left[7(r_2-y_2)\left(-\frac{dy_2}{dV_{23}}\right)\right]$$

$$= -\eta\left[7(r_2-y_2)\frac{d}{dV_{23}}\left(V^T Z_3+V_0\right)\right]$$

$$\Delta V_{23} = 7\eta(r_2-y_2)Z_3$$

ii)

$$\Delta W_{hj} = -\eta\frac{dE}{dW_{hj}}$$

$$= -\eta\sum \frac{dE^t}{dy^t}\frac{dy^t}{dZ_h^t}\frac{dZ_h^t}{dW_{hj}}$$

$$\frac{dZ_h^t}{dW_{hj}} = Z_h^t(1-Z_h^t)X_g^t$$

$$\frac{dy^t}{dZ_h^t} = V_h$$

$$\frac{dE^t}{dy^t} = \frac{d}{dy^t}\left[\frac{1}{2}\left[3(r_1-y_1)^2+7(r_2-y_2)^2\right]\right]$$

$$\frac{dE}{dy^1} = -3(r_1-y_1) \qquad \frac{dE}{dy^2} = -7(r_2-y_2)$$

$$\Delta W_{hj} = -\eta\left[\left(\frac{dE^1}{dy^1}\frac{dy^1}{dZ_h^1}\frac{dZ_h^1}{dW_{hj}}\right)+\left(\frac{dE^2}{dy^2}\frac{dy^2}{dZ_h^2}\frac{dZ_h^2}{dW_{hj}}\right)\right]$$

$$= -\eta\left[\left[-3(r_1-y_1)V_h Z_h^1(1-Z_h^1)\right] * X_g^1+\left[-7(r_2-y_2)V_h Z_h^2(1-Z_h^2)\right] * X_g^2\right]$$

$$= \eta\left[\left(3(r_1-y_1)V_h Z_h^1(1-Z_h^1)X_g^1\right)+\left(7(r_2-y_2)V_h Z_h^2(1-Z_h^2)X_g^2\right)\right]$$

3)
A)
Neural Net CB, Neural NetRZeroOne and NeuralNetCK cannot produce negative output
So only NeuralNetRK can produce negative outcome since it uses a linear function

B)
NeuralNetCK is the only one Neural Net which can ensure the sum of output to be 1.

C)
Neural NetCK is the best option here, since the sum of p1,p2,p3 which are the output is one. For this case where k=3 and p1,p2,p3 are the probabilities of class cat, face and tree which is similar to the output of neural net CK

D)
Neural NetRZeroOne is the best option here because of the following reasons,
i) The output is always 0 or 1
ii) All classification algorithms return probabilities we only need 0 or 1

4)
A)
We can explain this with an example, Suppose almond = 1, anise =2 , creostate = 3 and fishy = 4 then the neural net model would consider or treat fishy as a value which is double of anise which is not true, this can lead to a wrong answer, whereas in the case of random forest. Since neural network is an algorithm that learns from the weight of a variable its not ok.
For our example,
A simplified random forest algorithm can be,
If odor =1 do
Almond stuff
Else if odor ==2 do
Anise stuff........

Here there is no problem but when you consider an algorithm which learns by using weight as a parameter like
Prediction = weight * odor+.....
The weight matters, so its not correct to use label encoding but its useful to use one hot encoding.

B)
i)

Transformed Data Set (with optimal way of hot encoding)

	Z1	Z2	Z3	Z4	Z5	Label
X1	0	0	0	1	1	0
X2	0	0	1	0	0	0

Where z1 = 1 if odor = almond, z2 = 1 if odor = anise, z3 =1 of odor = fishy and z5 =1 if stalkshape= tapering.

Transformed Data Set (with hot encoding on every attribute)

	Z1	Z2	Z3	Z4	Z5	Z6	Label
X1	0	0	0	1	1	0	0
X2	0	0	1	0	0	1	0

ii) On using one hot encoding there is a problem of losing the hierarchy or order ie high > medium > low. So its better to use label encoding which retains the order.

iii) It is fine to use (0,1) instead of one hot encoding because it produces or represents the same dataset.
For example consider a dataset of coin flips

	Results
X1	Head
X2	Tail
X3	Head

On using just (1,0) with 1 = head and 0 = tail, we get

	Result
X1	1
X2	0
X3	1

And now applying one hot encoding we generate two variables namely d1 and d2 where d1 = 1 when result is head and d2 =1 when result is tail.

	D1	D2
X1	1	0
X2	0	1
X3	1	0

Here the column d2 is redundant and can be eliminated because only the column d1 is enough to represent the dataset. Since X1 and X3 have the same value and only X2 is different its is mandatory for X2 to have the value which is not X1 or X3. But when we consider three valued attributes the value of X2 can be either of the remaining two values not chosen by X1 or X3. So in the case of three valued attributes one hot encoding is compulsory.