EEE 591 Machine Learning with deployment to FPGA

HOMEWORK 3

ASU ID: 1225713099

Praveen Paidi

1.

$F(w) = a + b^T w + w^T C w$

$a=20$

$b^T = [1 \ -1]$

$C = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$

For a function to be convex,

Conditions are:

If for all pairs of functions, the line connects the two points passes above the curve.

Line connecting any two points in a set is lying within the set, it is said to be convex.

Second derivative should be positive:

$F'(w) = b^T + 2.C.w$

$F''(w) = 2.C$

$2.C = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$

Which is completely positive , it is convex.

2.b, 2.c **Fixed learning rate**

```
Fixed Learning Rate with Analytical Gradient
Gradient Descents iter  81
delay 0  ms time/iteration 0
learning rate alpha 0.1
The final weights: [[-0.5 ]
 [ 0.25]]


Fixed Learning Rate with Autograd
Gradient Descents iter  81
delay 0  ms time/iteration 0
learning rate alpha 0.1
The final weights: [[-0.5 ]
 [ 0.25]]
```

3. **Learning rate decrement**

```
Learning Rate decreases with Iteration by Analytical Gradient
Gradient Descents iter  27
delay 0  ms time/iteration 0
learning rate alpha 7.45e-09
The final weights: [[-0.44224238]
 [ 0.25      ]]


Learning Rate decreases with Iteration by Autograd
Gradient Descents iter  27
delay 0  ms time/iteration 0
learning rate alpha 7.45e-09
The final weights: [[-0.44224238]
 [ 0.25      ]]
```

4. **Normalized methods**

```
Normalized_methods with Iteration by Analytical Gradient
Gradient Descents iter  2001
delay 46.9  ms time/iteration 0.0234
learning rate alpha 0.5
The final weights: [[-0.5       ]
 [ 0.01422389]]


Normalized_methods with Iteration by Autograd
Gradient Descents iter  5001
delay 1.56e+03  ms time/iteration 0.312
learning rate alpha 0.5
The final weights: [[-0.5       ]
 [ 0.01422389]]
```

5. **Lipschitz Gradient**

```
Lipschitz with Iteration by Analytical Gradient
Gradient Descents iter  78
delay 0  ms time/iteration 0
learning rate alpha 0.5
The final weights: [[-0.5 ]
 [ 0.25]]


Lipschitz with Iteration by Autograd
Gradient Descents iter  78
delay 15.6  ms time/iteration 0.2
learning rate alpha 0.5
The final weights: [[-0.5 ]
 [ 0.25]]
```

6. **Steepest Descents**

```
Steepest Descents with Iteration by Analytical Gradient
Gradient Descents iter  2001
delay 172  ms time/iteration 0.0859
learning rate alpha 3.5e-06
The final weights: [[-0.35084424]
 [ 0.14915576]]


Steepest Descents with Iteration by Autograd
Gradient Descents iter  2001
delay 828  ms time/iteration 0.414
learning rate alpha 3.5e-06
The final weights: [[-0.35084424]
 [ 0.14915576]]
```

7. **Momentum with Steepest Descents**

```
Steepest momentum Descents with Iteration by Analytical Gradient
Gradient Descents iter  2001
delay 188  ms time/iteration 0.0937
learning rate alpha 3.52
The final weights: [[-0.5       ]
 [ 0.24983137]]


Steepest momentum Descents with Iteration by Autograd
Gradient Descents iter  2001
delay 875  ms time/iteration 0.437
learning rate alpha 3.52
The final weights: [[-0.5       ]
 [ 0.24983137]]
```

8. **With Regularizer**

```
Fixed Learning Rate with Analytical Gradient with regularizer
Gradient Descents iter  48
delay 0  ms time/iteration 0
learning rate alpha 0.1
The final weights: [[-0.33333333]
 [ 0.2       ]]


Fixed Learning Rate with Autograd with regularizer
Gradient Descents iter  48
delay 0  ms time/iteration 0
learning rate alpha 0.1
The final weights: [[-0.33333333]
 [ 0.2       ]]
```

Results appending in the list model results:

dict_classifiers	Totallearningrate	Totaldelay	Totalfinalweights
Fixed Learning Rate with Analytical Gradient	0.1000	0.0000	[[-0.5] [0.25]]
Fixed Learning Rate with Autograd	0.1000	0.0000	[[-0.5] [0.25]]
Learning Rate decreases with Iteration by Analytical Gradient	0.0000	0.0000	[[-0.44224238] [0.25]]
Learning Rate decreases with Iteration by Autograd	0.0000	0.0000	[[-0.44224238] [0.25]]
Normalized_methods with Iteration by Analytical Gradient	0.5000	46.8750	[[-0.5] [0.01422389]]
Normalized_methods with Iteration by Autograd	0.5000	1562.5000	[[-0.5] [0.01422389]]
Lipschitz with Iteration by Analytical Gradient	0.5000	0.0000	[[-0.5] [0.25]]
Lipschitz with Iteration by Autograd	0.5000	15.6250	[[-0.5] [0.25]]
Steepest Descents with Iteration by Analytical Gradient	0.0000	171.8750	[[-0.35084424] [0.14915576]]
Steepest Descents with Iteration by Autograd	0.0000	828.1250	[[-0.35084424] [0.14915576]]
Steepest momentum Descents with Iteration by Analytical Gradient	3.5185	187.5000	[[-0.5] [0.24983137]]
Steepest momentum Descents with Iteration by Autograd	3.5185	875.0000	[[-0.5] [0.24983137]]

The time taken by analytical gradient is lower than the autograd, it can be useful heavy computations rather than the autograd.
The weights are also different in different sections as the gradient proceeding way is different and internal function acting according to that.