Spring 2021
CMPE 462: Machine Learning
Assignment 2 Report
Sertay Akpinar - 2016400075

Part 1:

a) Step 1: Full Batch GD

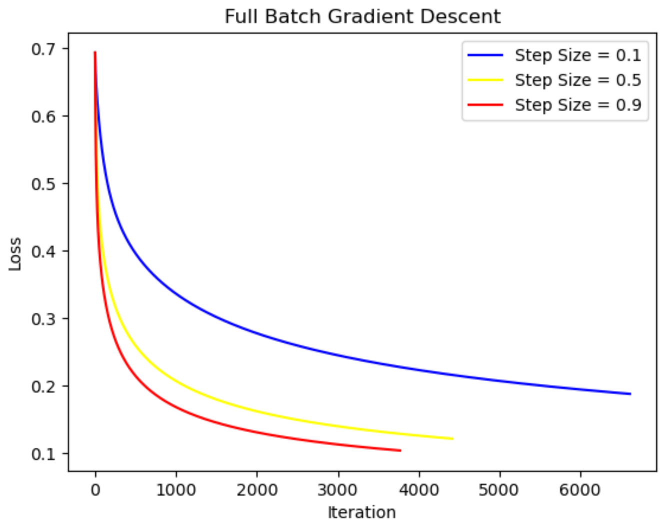

Figure a: Convergence Graph

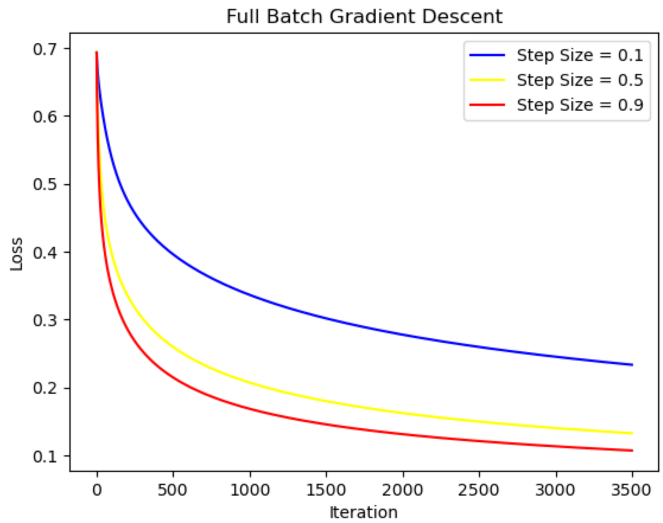

Figure b: 3500 Iterations

From the figure a we can observe that convergence speed is increasing from step size value of 0.1 to 0.9. That's why the number of iterations and time is decreasing from 0.1 to 0.9.

From the figure b and a we can observe that the remaining loss value is decreasing from step size value of 0.1 to 0.9. This means that the accuracy 0.9 is bigger than the accuracy of 0.1 (can be seen from the table below).

Step Size	Accuracy(≈)	Iterations	Time
0.1	0.949	6612	61.70
0.5	0.968	4416	40.87
0.9	0.973	3770	37.14

b) Step 2: Mini Batch SGD

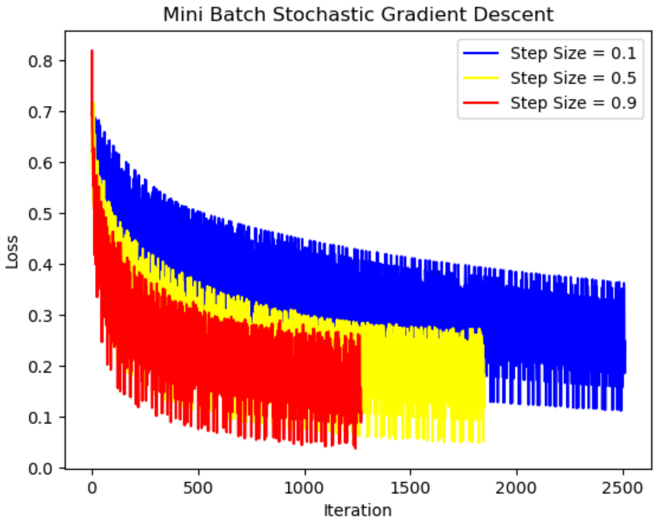

Figure c: Convergence Graph

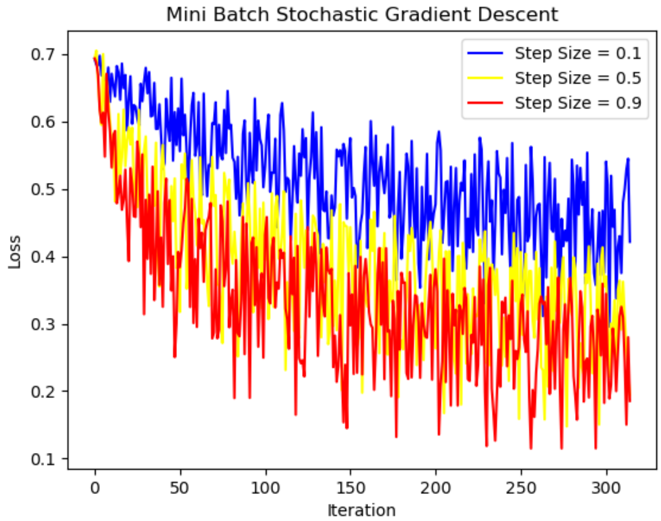

Figure d: 15 Epochs

From the figure c we can observe that the convergence speed is increasing from step size value of 0.1 to 0.9 also the number of iterations and time is decreasing from 0.1 to 0.9 as in the full batch GD. However the remaining loss value of 0.5 and 0.9 is nearly the same, since their accuracy results are the same based on rounding with 3 significant figures. The values can be changed in each run, since the algorithm contains **random sampling**.

From the figure d we can observe the fluctuations of mini batch stochastic GD clearly as stated in the class.

Step Size	Accuracy(≈)	Epoch Iterations	Time
0.1	0.927	121	0.92
0.5	0.959	90	0.90
0.9	0.959	62	0.50

The mini batch SGD converges faster than the full batch GD when we compare the time values from the tables. While the full batch GD has a smooth graph, the mini batch SGD's graph has fluctuations.

In both method's graphs, the following order from top to bottom is protected:

- Top: Blue → n = 0.1
- Middle: Yellow → n = 0.5
- Bottom: Red → n = 0.9

Notes:

- Table values are based on the convergence graphs.
- In mini batch SGD,
  → the iteration displayed in the x axis of the graph is different from the 'epoch iterations' stated in the table. In the graph, batch iterations are included.
  → the batch size is defined as 20.

Part 2:

What is asked?
$P(mammals \mid yes,\ no,\ yes,\ no)$
$P(non - mammals \mid yes,\ no,\ yes,\ no)$
We will use Bayes' rule:

$$P(Class \mid GiveBirth,\ CanFly,\ LiveInWater,\ HaveLegs) \propto P(GiveBirth,\ CanFly,\ LiveInWater,\ HaveLegs \mid Class)\ .\ P(Class)$$

From the given table in the assignment description, we can compute
$P(GiveBirth(yes) \mid mammals) = 6/7$
$P(GiveBirth(yes) \mid non - mammals) = 1/13$
$P(CanFly(no) \mid mammals) = 6/7$ $\qquad$ $P(CanFly(no) \mid non - mammals) = 10/13$
$P(LiveInWater(yes) \mid mammals) = 2/7$
$P(LiveInWater(yes) \mid non - mammals) = 3/13$
$P(HaveLegs(no) \mid mammals) = 2/7$
$P(HaveLegs(no) \mid non - mammals) = 4/13$

Let's compute the posterior probabilities:

$$P(mammals \mid GiveBirth(yes)\ ,\ CanFly(no),\ LiveInWater(yes),\ HaveLegs(no)) \propto$$
$$P(GiveBirth(yes) \mid mammals) \cdot P(CanFly(no) \mid mammals) \cdot P(LiveInWater(yes) \mid mammals) \cdot$$
$$P(HaveLegs(no) \mid mammals) \cdot P(mammals) = 6/7 \cdot 6/7 \cdot 2/7 \cdot 2/7 \cdot 7/20 \approx 0.021$$

and

$$P(non - mammals \mid GiveBirth(yes)\ ,\ CanFly(no),\ LiveInWater(yes),\ HaveLegs(no)) \propto$$
$$P(GiveBirth(yes) \mid non - mammals) \cdot P(CanFly(no) \mid non - mammals) \cdot$$
$$P(LiveInWater(yes) \mid non - mammals) \cdot P(HaveLegs(no) \mid non - mammals) \cdot P(non - mamm$$
$$= 1/13 \cdot 10/13 \cdot 3/13 \cdot 4/13 \cdot 13/20 \approx 0.003$$

Decision:

$$0.021 > 0.003$$
$$P(mammals \mid GiveBirth(yes)\ ,\ CanFly(no),\ LiveInWater(yes),\ HaveLegs(no))$$
$$>$$
$$P(non - mammals \mid GiveBirth(yes)\ ,\ CanFly(no),\ LiveInWater(yes),\ HaveLegs(no))$$

As a result, the test sample is predicted as mammals.