Indian Institute of Technology Delhi

COL774 - Machine Learning

Assignment 1

Linear and Logistic regression

Garv Nagori
2021CS10549

Contents

§1 Question 1

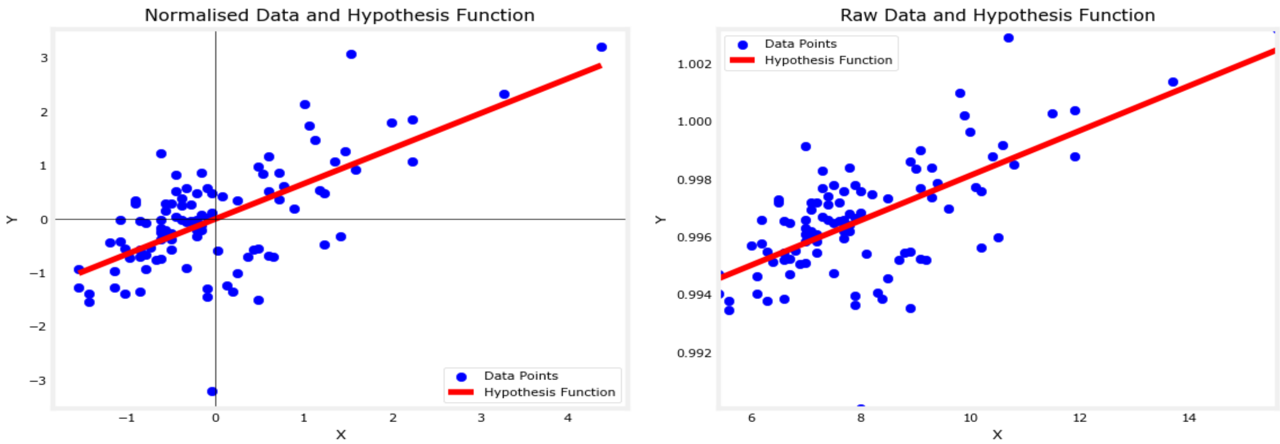
§1.1 Subpart a.

The learning rate used was $\eta = 0.01$.
The stopping criteria was if the difference in absolute error was $< 10^{-11}$. The final set of parameters were:

$$\theta = [-5.99244196 \times 10^{-14} 6.55045094 \times 10^{-1}]$$
$$\bar{} = [0.9959931, 7.7777112\ times 10^{-5}]$$

where θ is the parameters for normalised input and $\bar{\theta}$ is the parameters for unnormalised input.

§1.2 Subpart b.



(a) Normalised Data

(b) Raw Data

Figure 1: The data and the hypothesis function

§1.3 Subpart c.

The animations are in the directory.

§1.4 Subpart d.

The animations are in the directory.

§1.5 Subpart e.

The animations are in the directory.

§2 Question 2

§2.1 Subpart a.

The data points were sampled using np.random.normal

§2.2 Subpart b.

In each case, I observed the error for 10 iterations and checked if their average was less than some *conv_crit*, or the convergence criteria.
For lesser batch sizes, it was 0.1 but for bigger batch sizes it was closer to 1 (.95).

The parameters for different values of batch size are:

$$batchsize = 1 - [2.81446261.005667291.96551638]$$
$$batchsize = 100 - [2.737533151.052214411.98514639]$$
$$batchsize = 10000 - [2.8171141.04066131.98715]$$
$$batchsize = 1000000 - [2.8166551.0405671.9871567]$$

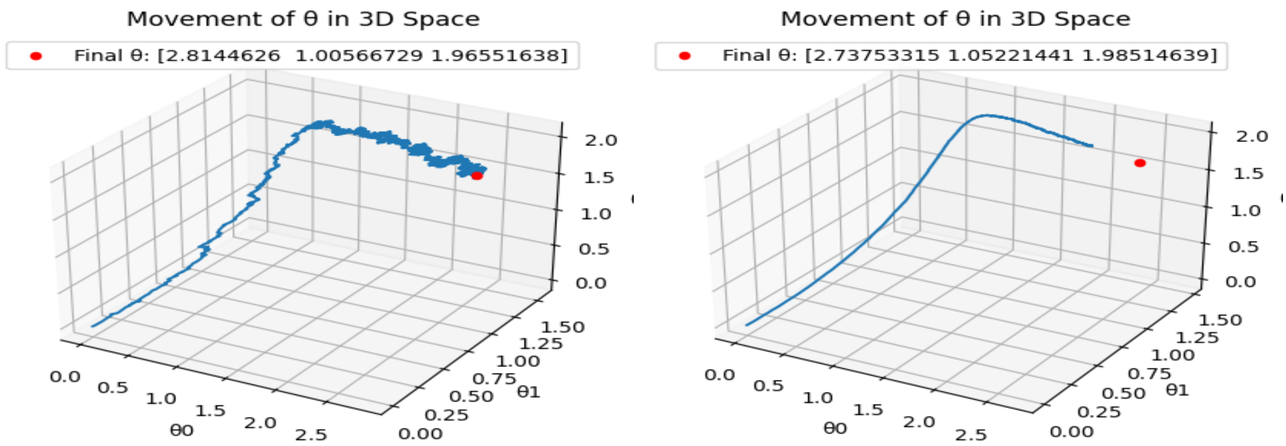
§2.3 Subpart c.

The speed of SGD is the highest whereas speed of Batch Gradient Descent is lowest.
The number of iteration first decreases and then increases as we increase the mini-batch size
The error values for them are:

$$actual - 0.9829467$$
$$batchsize = 1 - 1.06008464$$
$$batchsize = 100 - 1.14831872$$
$$batchsize = 10000 - 1.0819431$$
$$batchsize = 1000000 - 1.0815123$$

§2.4 Subpart d.

§2.5 Subpart b.



(a) Batch Size 1

(b) Batch size 100

Figure 2: The data and the hypothesis function

SGD is noisy since variance of movement is high

§3 Question 3

§3.1 Subpart a.

Newton's method clearly converged way faster since it converged in 10 iterations in my case.
The final parameter values obtained are:

$$\theta = [0.40125316, 2.5885477, -2.7255884]$$

§3.2 Subpart b.

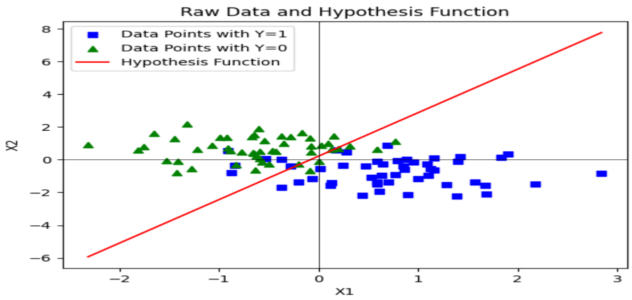


Figure 3: The Data Points and the Decision Boundary

§4 Question 4

§4.1 Subpart a.

The values of the different parameters are given by

$$\begin{aligned}\mu_0 &= [0.75529433, -0.68509431]\\ \mu_1 &= [-0.75529433, 0.68509431]\\ \sigma &= [[0.42953048, -0.02247228],\\ &[-0.02247228, 0.53064579]]\end{aligned}$$

§4.2 Subpart b.

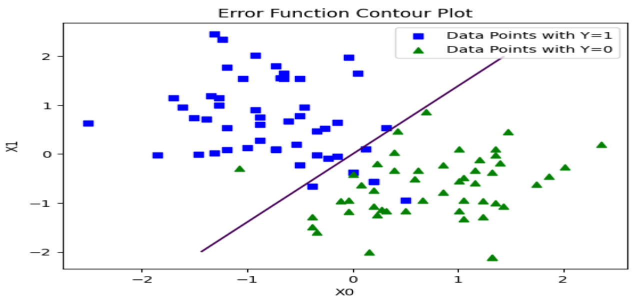


Figure 4: The Decision Boundary for Linear GDA

§4.3 Subpart c.

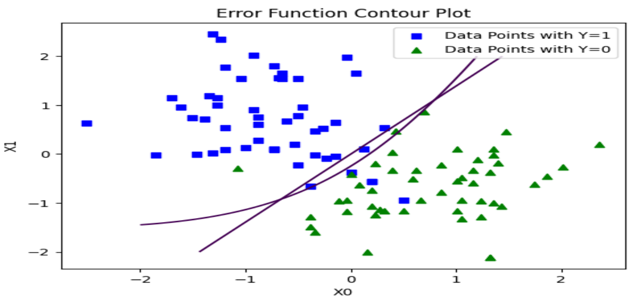
The equation of linear boundary:

$$\theta^T x = 0 \ \ where \ \ \theta = [\theta_0, \theta_1, \theta_2]$$

$$\theta_0 = \mu_1^T \Sigma^{-1} \mu_1 - \mu_0^T \Sigma^{-1} \mu_0$$

and,

$$[\theta_1, \theta_2] = 2 * \Sigma^{-1}(\mu_0 - \mu_1)^T$$



H

Figure 5: The Decision Boundary for Quadratic GDA

§4.4 Subpart d.

The values of the parameters are as follows:

$$\begin{aligned}
\mu_0 &= [0.75529433, -0.68509431] \\
\mu_1 &= [-0.75529433, 0.68509431] \\
\sigma_0 &= [[0.47747117, 0.1099206], \\
&\quad [0.1099206, 0.41355441]] \\
\sigma_1 &= [[0.38158978, -0.15486516], \\
&\quad [0.1099206, 0.41355441]]
\end{aligned}$$

§4.5 Subpart e.

The equation for the quadratic boundary is given by,

$$x^T(\Sigma_1^{-1} - \Sigma_0^{-1})x + 2 * x^T(\Sigma_0^{-1}\mu_0 - \Sigma_1^{-1}\mu_1) + (\mu_1^T\Sigma_1^{-1}\mu_1 - \mu_0^T\Sigma_0^{-1}\mu_0) = 0$$

§4.6 Subpart f.

The quadratic equation gives a better decision boundary than the linear.

§5 Acknowledgements

We have used the style file from here[1] to produce this document.

[1] https://github.com/vEnhance/dotfiles/blob/main/texmf/tex/latex/evan/evan.sty