Machine Learning: Assignment 1

Gohil Dwijesh
2017CS50407

January 2020

1 Linear Regression

- Learning rate: 0.0001.
- stopping criteria: if modulus of the derivative of the cost function is less than 0.0005.
- Matrix formula used:

$$\frac{\partial J(\theta)}{\partial \theta} = \frac{1}{m} * (X^{\mathrm{T}} X\theta - X^{\mathrm{T}} Y)$$

$$J(\theta) = \frac{1}{2m}(Y - X\theta)^{\mathrm{T}}(Y - X\theta)$$

- Where X is an (m x n) matrix, m = number of training examples, and n = number of features.

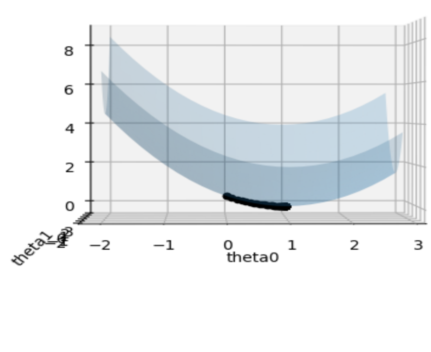

Figure 1: J(θ) as a function of θ_0, θ_1

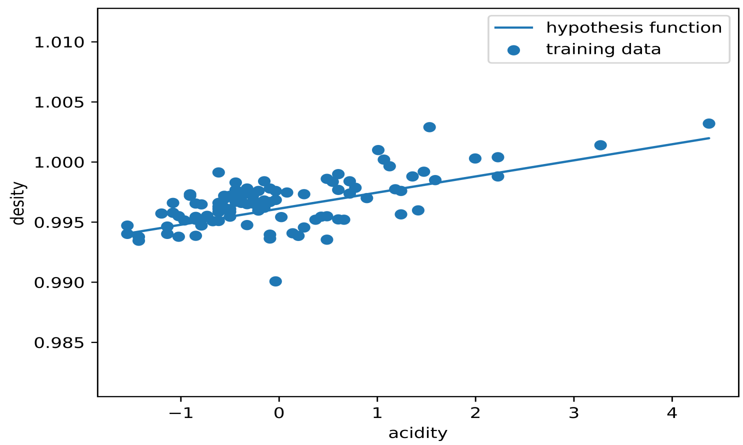

Figure 2: h_θ(x) and training data

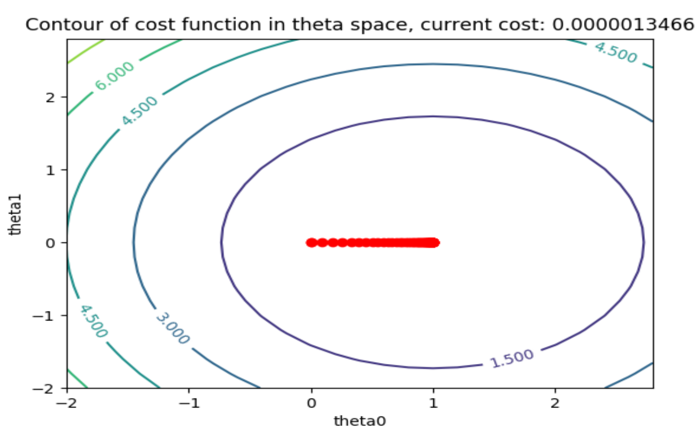

Figure 3: Question 1: Contour of cost function and iteration points

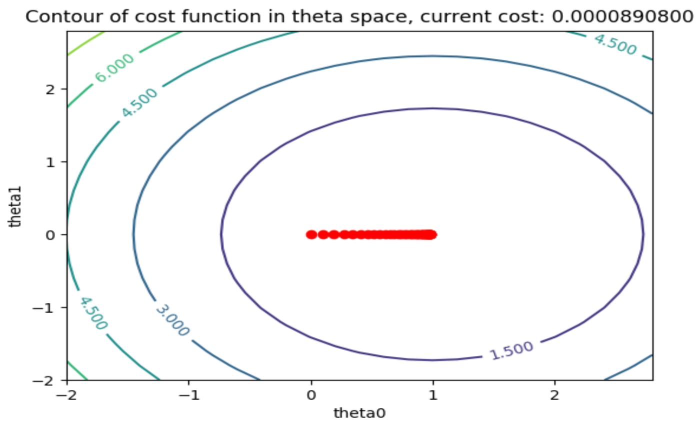

Figure 4: Question 1: Contour of cost function at learning rate: 0.1

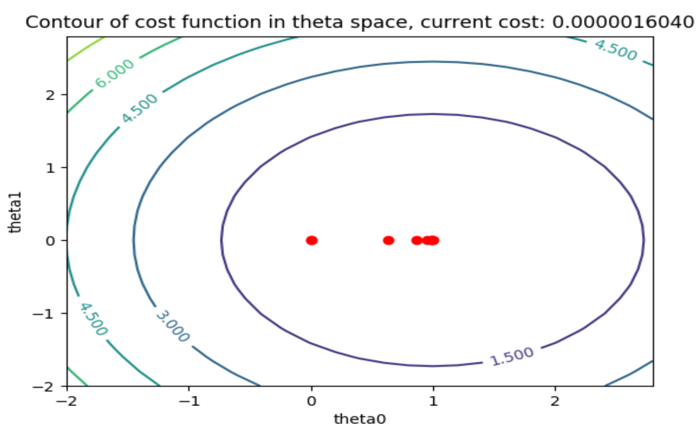

Figure 5: Question 1: Contour of cost function at learning rate: 0.001

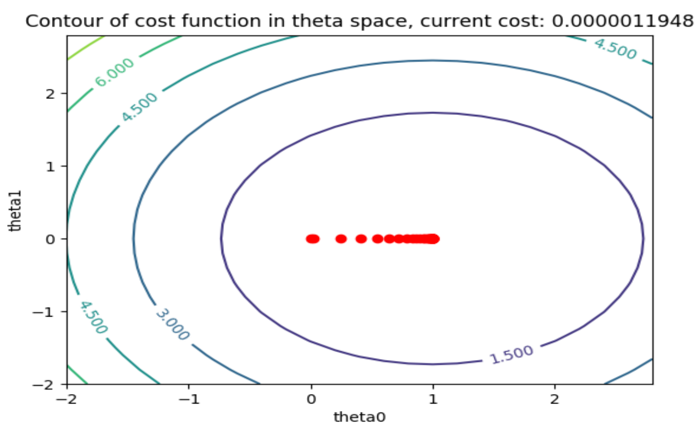

Figure 6: Question 1: Contour of cost function at learning rate: 0.025

2 Sampling and Stochastic Gradient Descent

2.1 Implementing stochastic gradient descent

Batch size	$theta_0$	$theta_1$	$theta_2$	epochs	opt. avg	history	error
1	2.98841157	0.98489459	1.95629993	1	NA	5000	1.09743707
100	2.99857972	1.00004456	1.99967958	5	0.531	2000	0.98300276
10000	2.91204599	1.01900053	1.99320696	126	0.05	50	1.00538196
1000000	3.0002548	1.0002631	1.9995629	23753	0.0012	1	0.98300278
NA	3	1	2	NA	NA	NA	0.98294692

Table 1: Sampling and Stochastic Gradient Descent

2.2 Explanation

- As the batch size increases the number of epochs before converging increases.
- As the batch size decreases the speed of convergence increases.
- As the batch size increases, it converges to assumed hypothesis theta values, because it is going towards batch gradient descent as we increase the batch size.

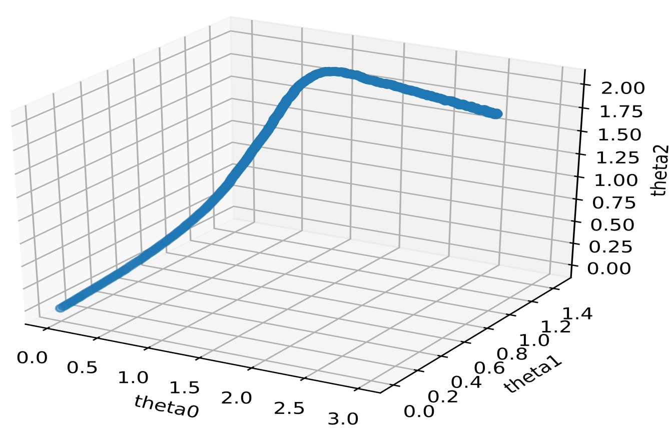

Figure 7: Question 2: batch size: 100

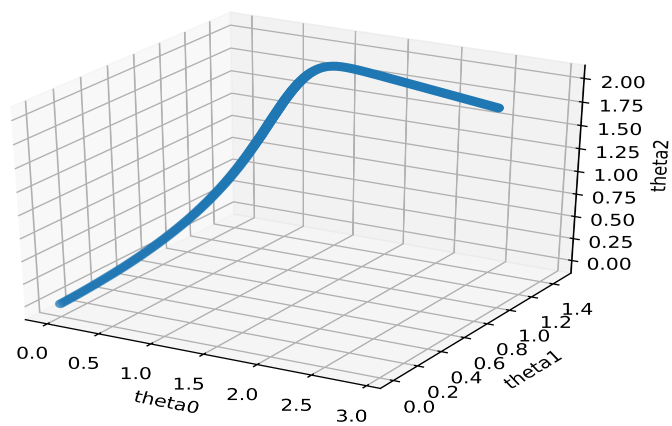

Figure 8: Question 2: Batch size: 10000

3 Logistic Regression

- θ: 0.40125316, 2.5885477, -2.72558849

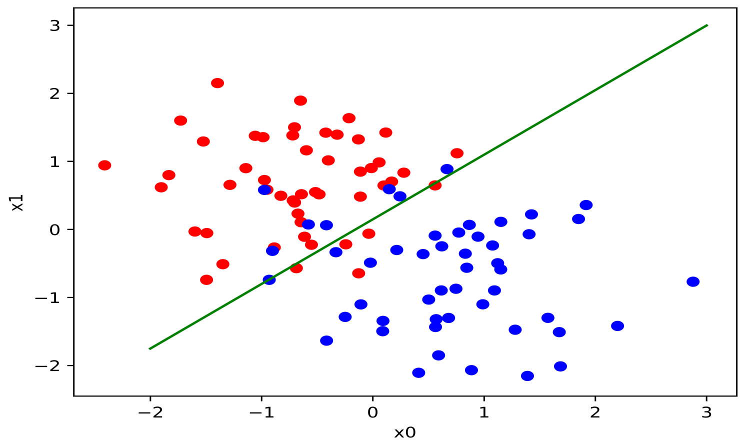

Figure 9: Question 3: Linear boundary separating the two classes

4 Gaussian Discriminant Analysis

0.5

Table 2: ϕ

-0.75529433
0.68509431

Table 3: μ_0

0.75529433
-0.68509431

Table 4: μ_1

0.38158978	-0.15486516
-0.15486516	0.64773717

Table 5: Σ_0

0.47747117	0.1099206
0.1099206	0.41355441

Table 6: Σ_1

0.42953048	-0.02247228
-0.02247228	0.53064579

Table 7: Σ

- Equation of the boundary separating the two classes:

  $$x.^{\mathrm{T}}({\Sigma_1}^{-1}-{\Sigma_0}^{-1})x+2({\mu_0}^{\mathrm{T}}{\Sigma_0}^{-1}-{\mu_1}^{\mathrm{T}}{\Sigma_1}^{-1})x+{\mu_1}^{\mathrm{T}}{\Sigma_1}^{-1}\mu_1-{\mu_0}^{\mathrm{T}}{\Sigma_0}^{-1}\mu_0+\log_e((\frac{1-\phi}{\phi})^2\frac{|\Sigma_1|}{|\Sigma_0|})=0$$

- If we assume that our features shares the Σ matrix then we get a linear boundary and if we assume that the features can have different Σś then we get a quadratic boundary. Looking at the training data, assuming the later case is a good choice.

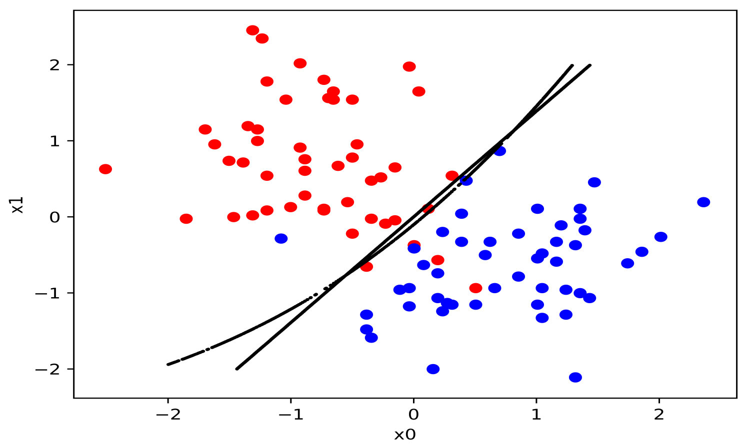

Figure 10: Question 4: Boundary separating the two classes