



دانشگاه یزد

گروه مهندسی کامپیوتر

رشته تحصیلی: هوش مصنوعی و رباتیکز

نام درس: یادگیری ماشین

تکلیف شماره ۱

استاد مربوطه: دکتر مهدی یزدیان

تهیه کننده: حمیدرضا نادمی

بهار ۱۳۹۸

### A. Linear Regression with one variable

I. Consider the attached file dataset1.txt. The first column of the data file shows the input data (x) and the second column shows the output value (y) for each sample.

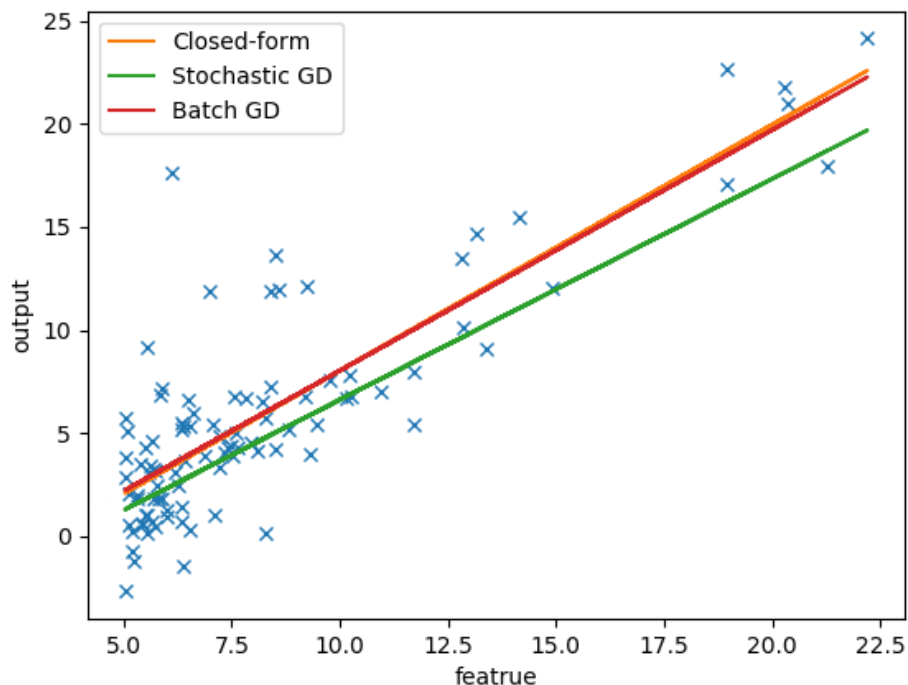
- 1) Fit a linear regression model on your data using :
  - a) Closed-form solution calculated by LSE method
  - b) Gradient descent method in online (stochastic) mode (1500 iterations)
  - c) Gradient descent method in batch mode (1500 iterations)

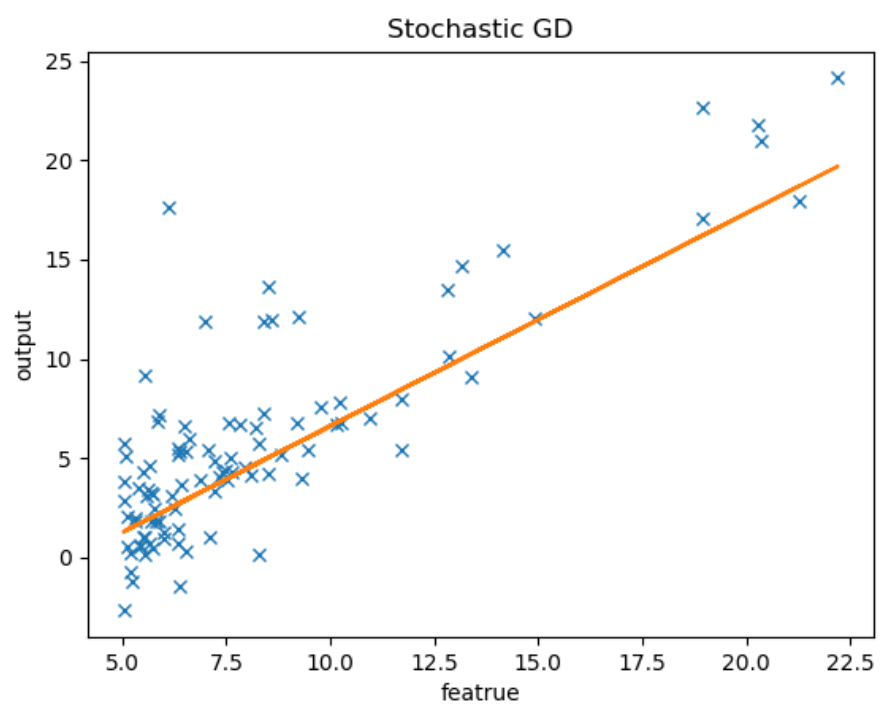
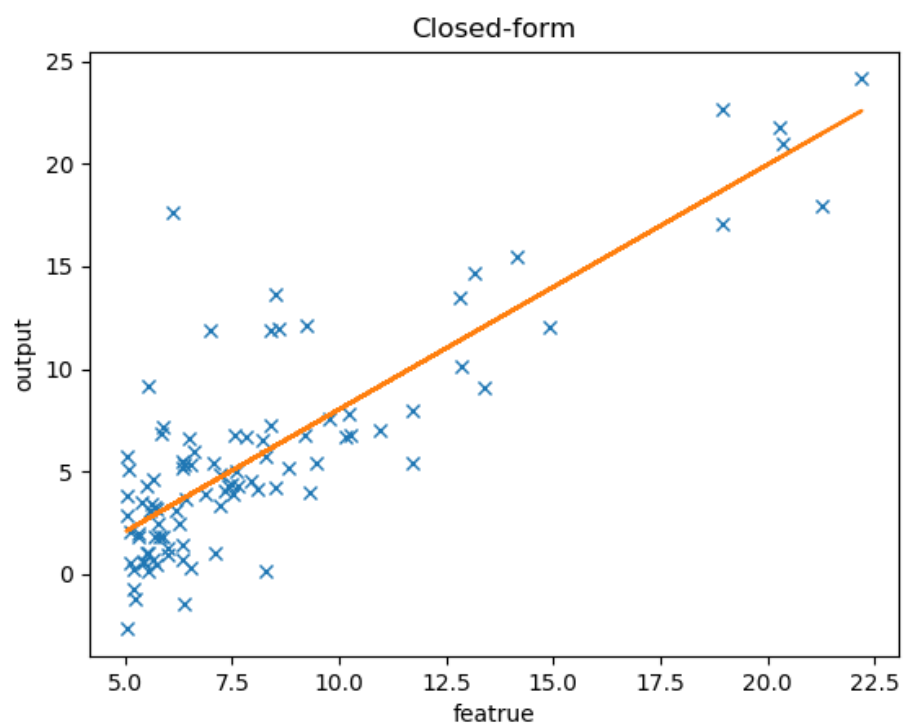
	$\theta_0$	$\theta_1$
<b>Closed-form</b>	-3.8957	1.1930
<b>Stochastic GD</b>	-4.0969	1.0717
<b>Batch GD</b>	-3.6403	1.1673

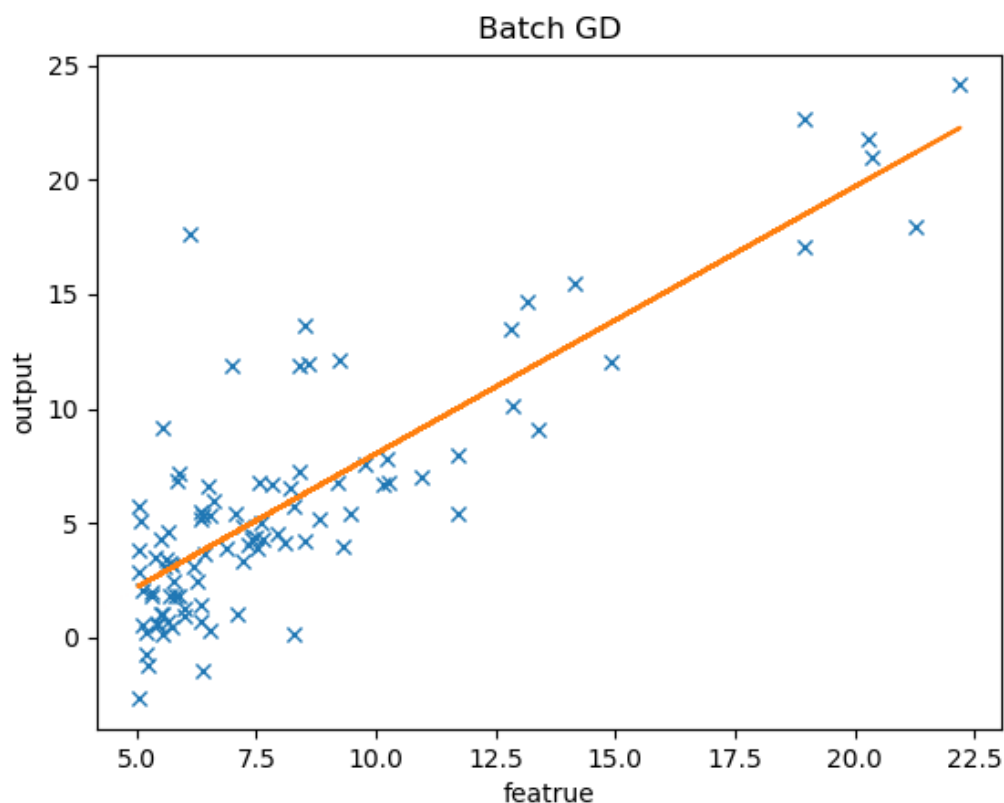
- 2) What is the equation of cost function  $J(\theta)$  for linear regression?

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

- 3) Plot the dataset and superimpose the fitted models using three above methods.







4) Use each estimated parameter  $\hat{\theta}$  to predict the output for  $x=6.2, 12.8, 22.1, 30$ .

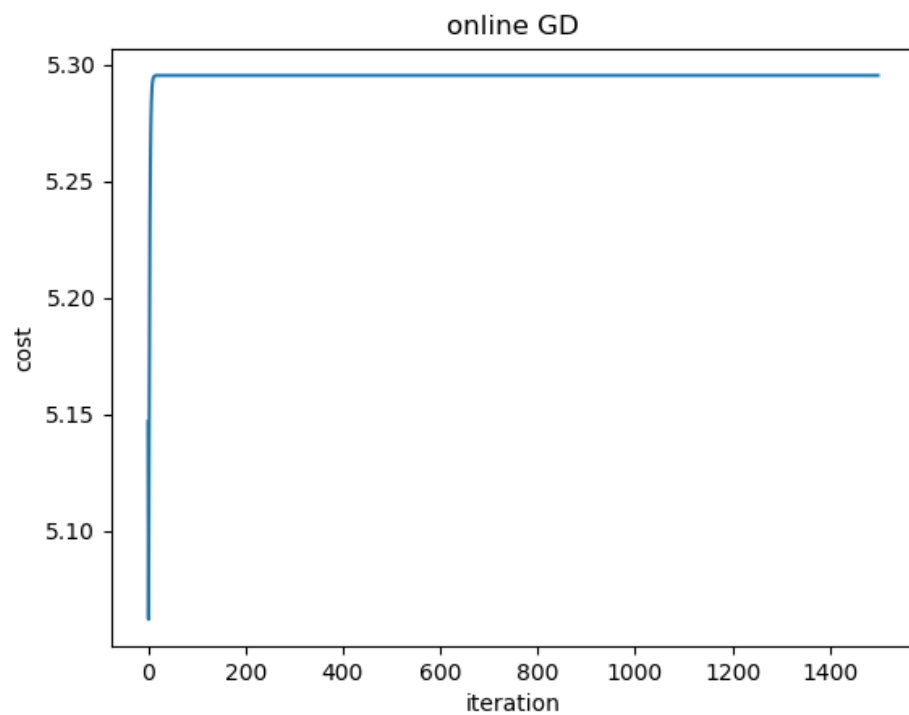
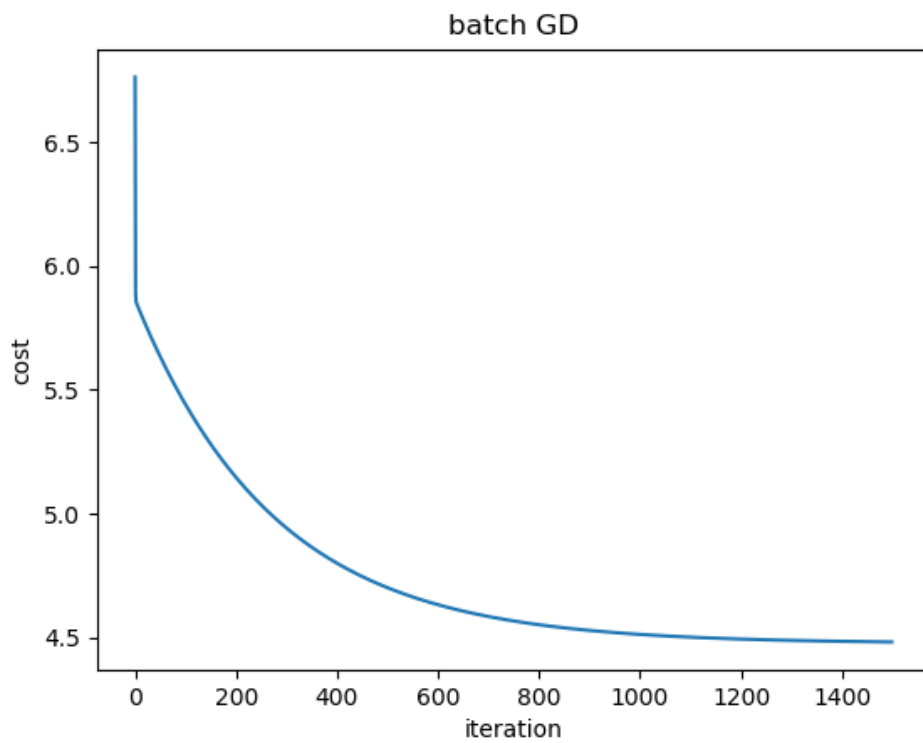
	Closed-form model	Stochastic GD model	Batch GD model
<b>X=6.2</b>	3.5010	2.5477	3.5973
<b>X=12.8</b>	11.3750	9.6210	11.3020
<b>X=22.1</b>	22.4702	19.5880	22.1585
<b>X=30</b>	31.8952	28.0546	31.3808

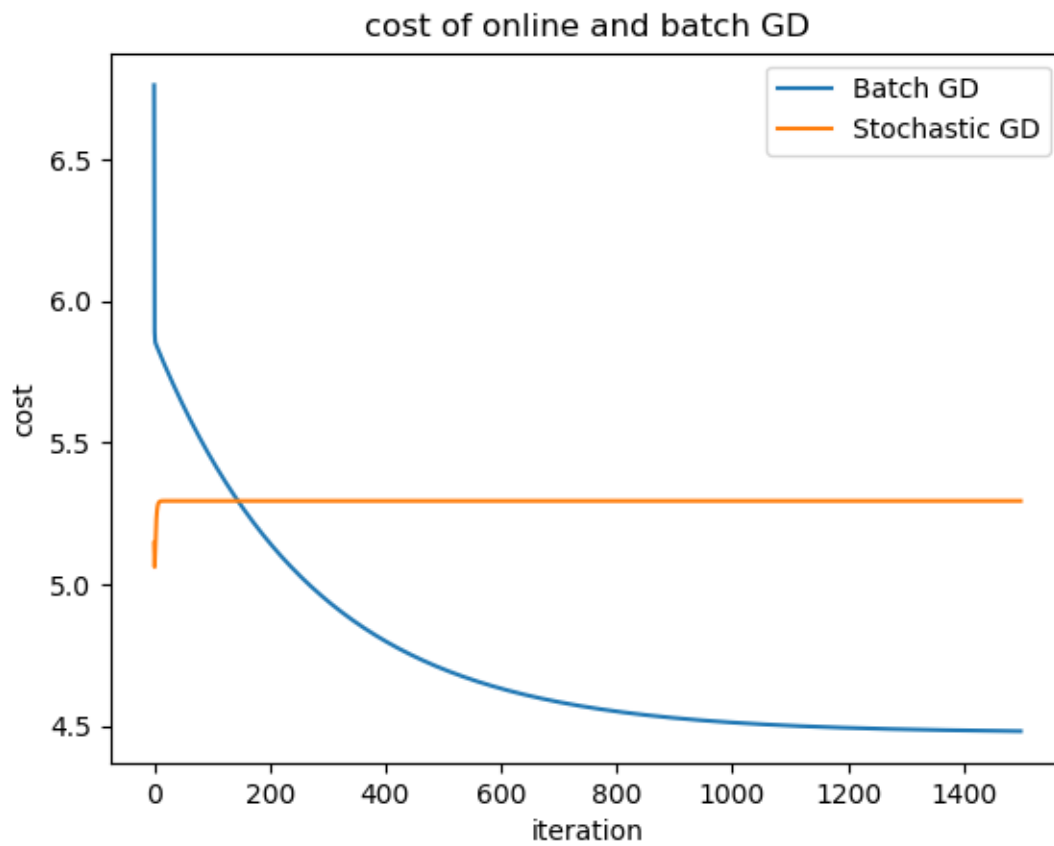
5) Compare the parameter  $\hat{\theta}$  estimated by each method.

با توجه به اینکه  $J(\theta)$  یک تابع convex است و فقط یک Global minimum دارد، روش Batch GD پارامترهای بهینه را پیدا میکند. در تخمین پارامتر به روش Closed-form نیز پارامترهای بهینه تخمین زده می شوند که مقدار ناچیزی با مقدار بدست آمده با روش Batch GD تفاوت دارد. در روش Stochastic GD مقدار گرادیان در هر گام تنها برای یک داده از دیتاست محاسبه می شود بنابراین دقت آن نسبت به پارامترهای تخمین زده شده با Batch GD کمتر است ولی سرعت رسیدن آن به پارامتر بهینه بیشتر است.

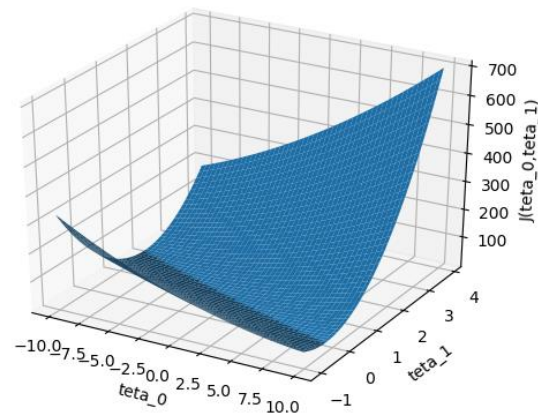
بهینه ترین پارامترها با روش Batch GD تخمین زده شده است.

- 6) Plot the cost function  $J(\theta)$  along the epochs (plot both online & batch methods on one figure using **hold on** command).





- 7) To understand the cost function better, plot  $J(\theta)$  in terms of  $\theta_0$  and  $\theta_1$ . (Note: generate a grid of  $\theta_0 \in [-10 \ 10]$  and  $\theta_1 \in [-1 \ 4]$  and plot the cost function,  $J(\theta_0, \theta_1)$ , using **surf** command).



8) Which type of G.D. (online\batch) do you prefer here? Why?

Batch GD, زیرا دقت آن بیشتر است و در این مسئله تعداد داده‌ها نیز کم هستند و مدت زمان پیدا کردن پارامترهای بهینه زیاد نمی شود.

### **B. Analysis the effect of outlier**

In this part we are going to analysis the effect of outlier on liner regression method. Here, use the dataset1 in the previous part.

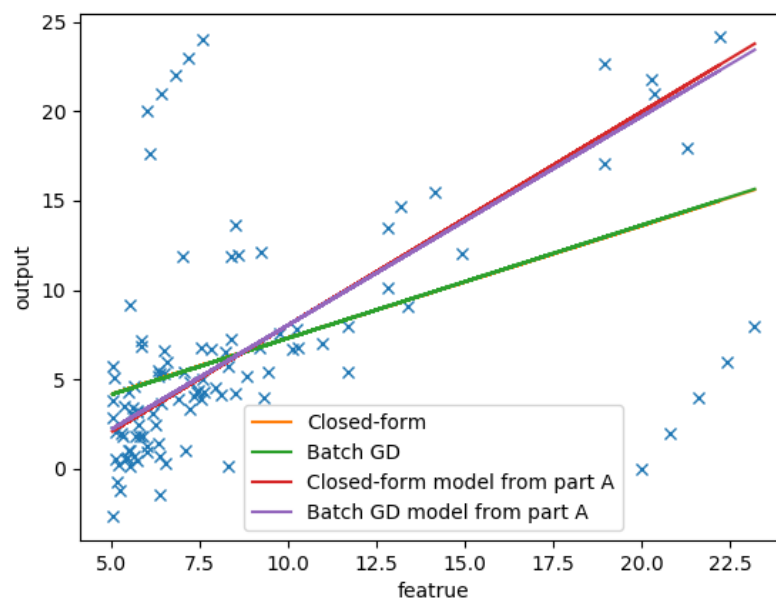
- 1) Add some outlier by randomly generating some samples in the border of the data range.  
For this, generate 5 samples with  $x \in [6 \ 8]$  and  $y \in [20 \ 25]$  and 5 samples with  $x \in [20 \ 24]$  and  $y \in [0 \ 10]$ .

فایل utility.py ، خط ۱۲۳ الی ۱۳۵

- 2) Fit a model on the new data using:
  - a) Closed form solution of liner regression.
  - b) Batch Gradient descent (1500 iterations).

	$\theta_0$	$\theta_1$
<b>Closed-form</b>	1.0226	0.6287
<b>Batch GD</b>	0.9862	0.6320

- 3) Plot the data and superimpose the fitted models from Part A (i.e., A-1-a & A-1-c) and in this part onto your figure.



4) Use the estimated parameter  $\hat{\theta}$  to predict the output for  $x=6.2, 12.8, 22.1, 30$ .

	Closed-form model	Batch GD model
<b>X=6.2</b>	4.9211	4.9048
<b>X=12.8</b>	9.0711	9.0763
<b>X=22.1</b>	14.9188	14.9543
<b>X=30</b>	19.8862	19.9475

5) How do you explain the effect of outlier on linear regression method?

Outlier باعث می شود تا مدل تخمین زده شده نسبت به داده های پرت کمی متمایل شود.

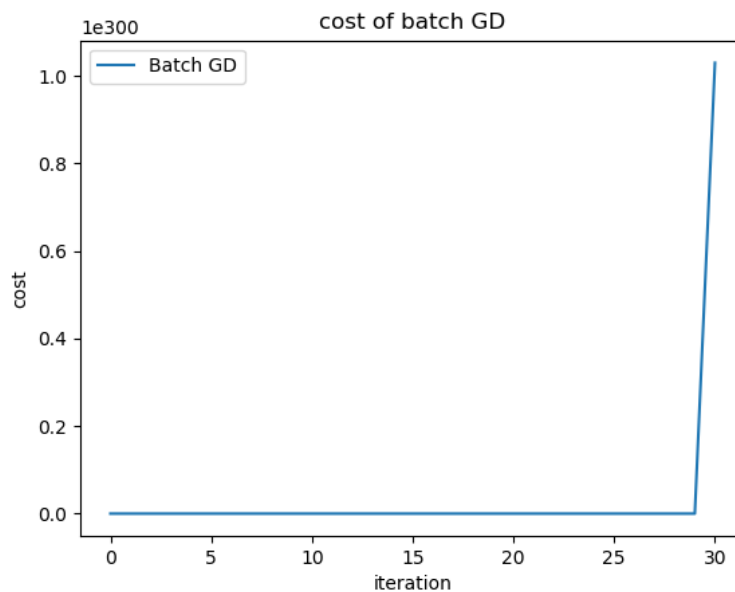
### C. Linear Regression with multiple variables

The data file 'dataset2.txt' contains a 2-dimensional data. The first two columns of the data file show the feature of each sample and the last column illustrate its corresponding output value.

- 1) Fit a linear regression model on your data using :
  - a) Closed form solution of liner regression.
  - b) Batch Gradient descent (1500 iterations).

	$\theta_0$	$\theta_1$	$\theta_2$
<b>Closed-form</b>	89597.9095	139.2106	-8738.0191
<b>Batch GD</b>	nan	nan	nan

2) Plot the cost function  $J(\theta)$  along the epochs for Gradient descent method.





- 3) Use each estimated parameter  $\hat{\theta}$  to predict the output for  $x = [1357, 5]$ ,  $x=[2500, 4]$ .

	Closed-form model	Batch GD model
<b>X=[1357,5]</b>	234816.6986	nan
<b>X=[2500,4]</b>	402672.5181	nan

- 4) Compare the parameter  $\hat{\theta}$  estimated by each method.

در تخمین پارامترها با روش Batch GD به دلیل نرمالایز نشدن پارامترها به نقطه همگرایی متمایل نشد و واگرا شد و مقدار nan را برای هر سه پارامتر تخمین زد.

#### D. Analysis the effect of feature renormalization

- 1) Apply feature normalization on the dataset2 used in Part C (Note: for normalization you can scale the samples such a way that the minimum and maximum values of each feature change to 0 and 1, respectively).

فایل utility.py ، خط ۱۱۳ الی ۱۲۵

- 2) Fit a model on the new data using
- Closed form solution of liner regression.
  - Batch Gradient descent (1500 iterations)

	$\theta_0$	$\theta_1$	$\theta_3$
<b>Closed-form</b>	340412.6595	109447.7964	-6578.3548
<b>Batch GD</b>	340412.5630	109370.0567	-6500.6150

- 3) Use the estimated parameter  $\hat{\theta}$  to predict the output for  $x = [1357, 5]$ ,  $x=[2500, 4]$ .

	Closed-form model	Batch GD model
<b>X=[1357,5]</b>	148828180.6946	148723076.4357
<b>X=[2500,4]</b>	273933590.4142	273739551.8642

- 4) Explain and Discuss about the results of the methods before and after feature normalization.

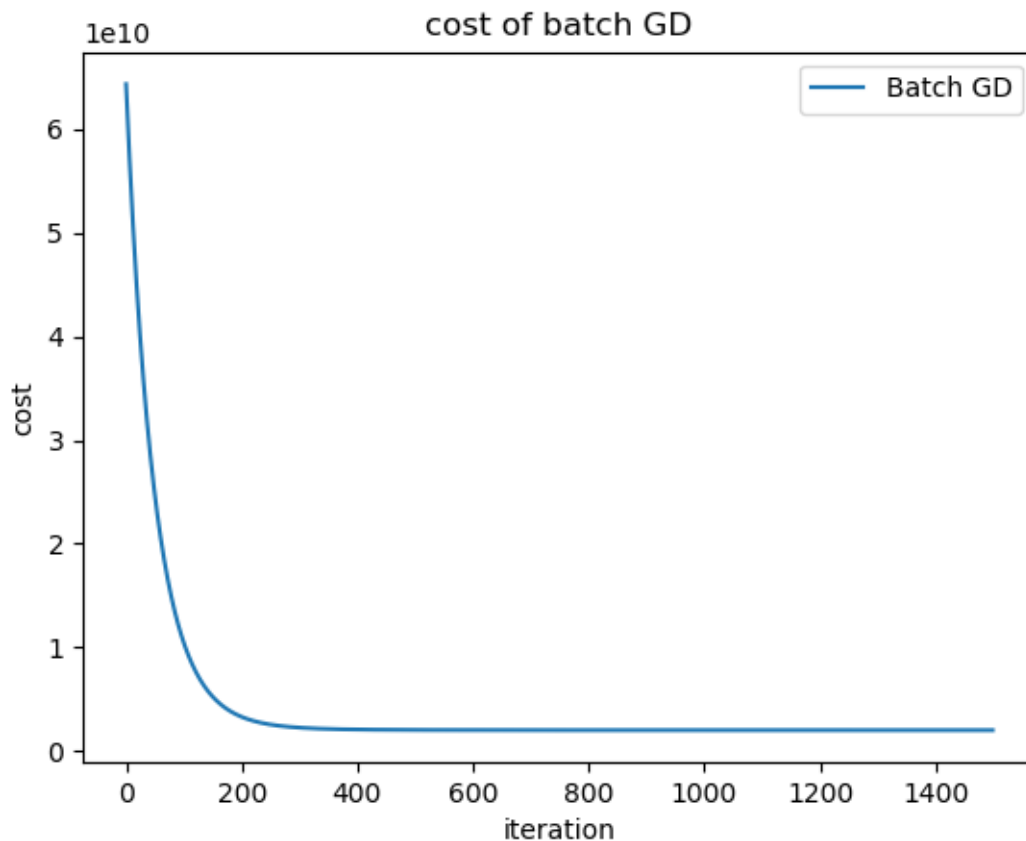
قبل از نرمال کردن ویژگی‌ها نمودار خطا در هر گام از iteration یک روند واگرا داشت و مقدار nan برای پارامترهای  $\theta_0$ ،  $\theta_1$ ،  $\theta_2$  بدست آمد که بعد از اعمال feature normalization روی داده‌ها و نرمال کردن

آنها با فرمول زیر با افزایش گام ها مقدار خطا کاهش یافت و روندی نزولی پیدا کرد و به پارامترهای  $\theta_0, \theta_1, \theta_2$  به بهینه ترین مقدار همگرا شدند.

$$\tilde{x}_i = \frac{x_i - \mu_i}{\sigma_i} \quad \forall \text{feature } x_i$$

*Mean Normalization*

با اعمال feature normalization سرعت اجرا Batch GD در مقایسه با قبل بیشتر شد.



### **E. Logistic Regression**

Consider the attached file dataset3.txt. The first two columns of the data file show the feature of each sample and the last column illustrates its corresponding binary level.

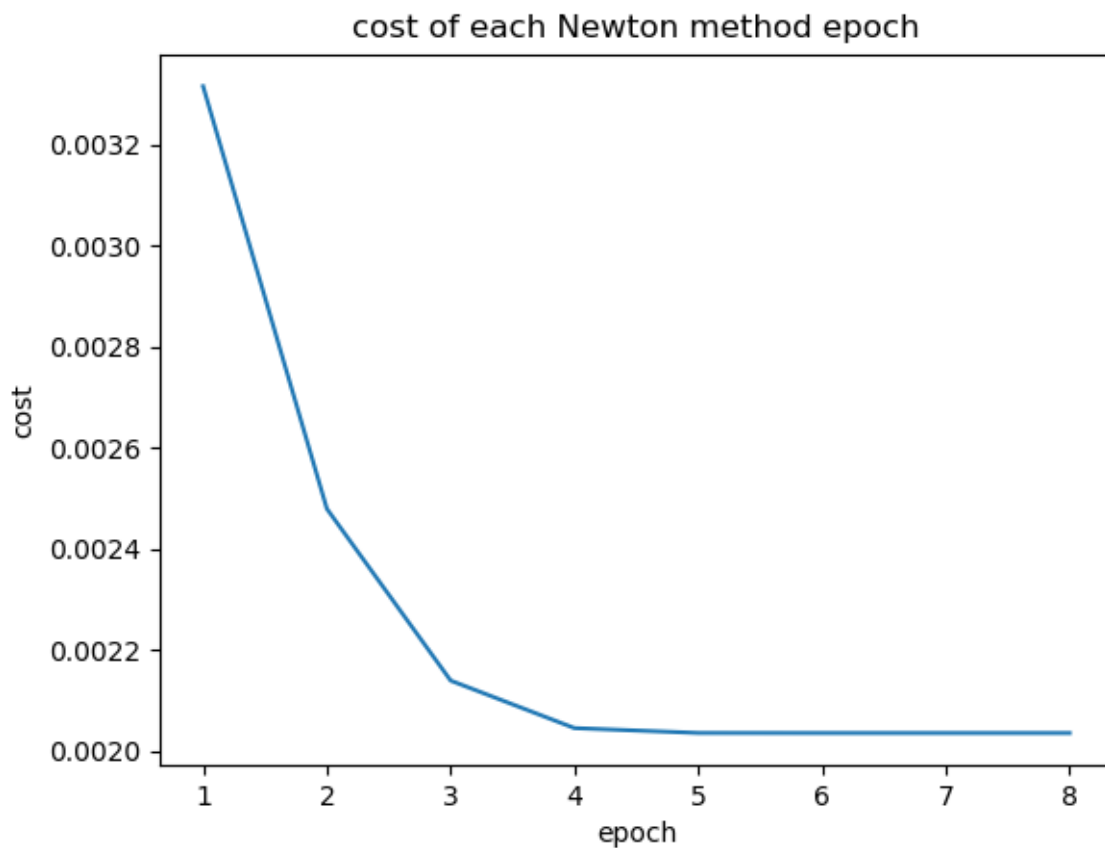
- 1) What is the cost function  $J(\theta)$  in logistic regression?

$$J(\theta) = -\frac{1}{m} \left[ \sum_{i=1}^m y^{(i)} \log(h_{\theta}(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_{\theta}(x))^{(i)} \right]$$

2) Estimate the parameter  $\hat{\theta}$  using Newton method.

	$\theta_0$ (bias)	$\theta_1$	$\theta_2$
<b>GD</b>	-25.1613	0.2062	0.2014

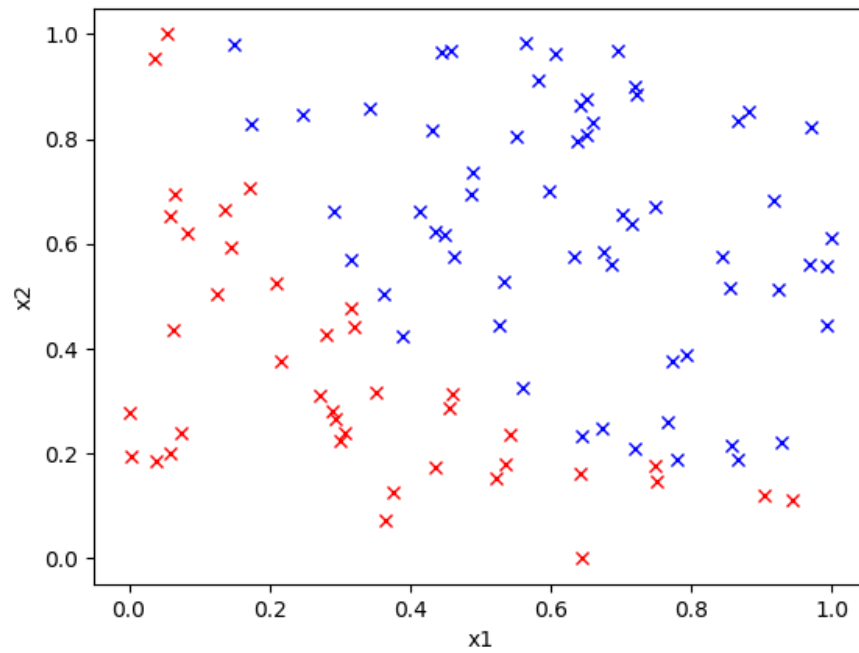
3) Plot the cost function  $J(\theta)$  along the epochs of the Newton method.



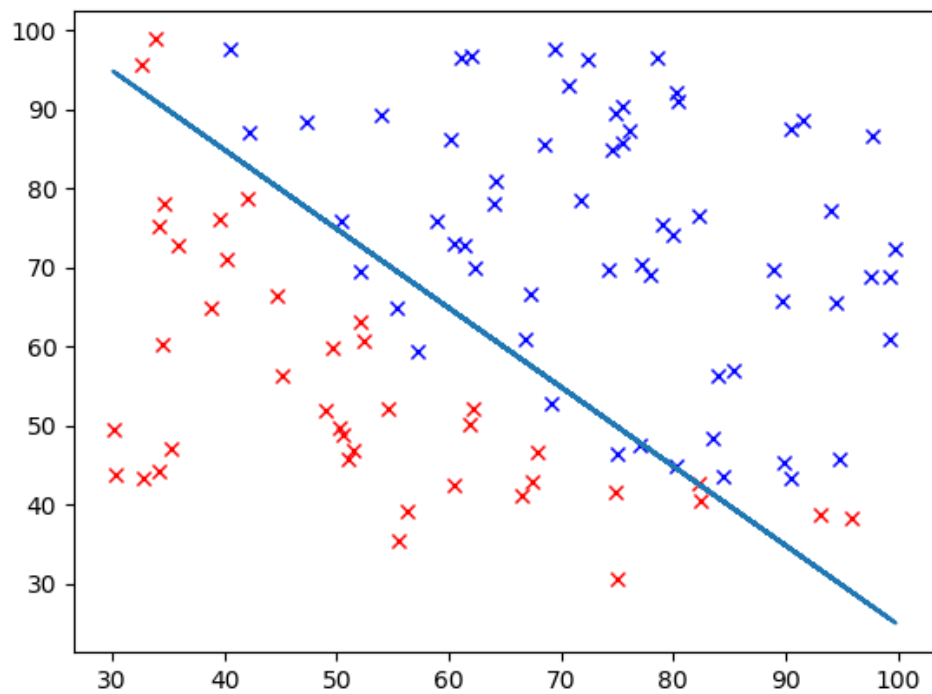
4) Use the learned model to classify all training example.

$$\text{Confusion matrix} = \begin{bmatrix} 34 & 6 \\ 5 & 55 \end{bmatrix}$$

5) Plot the data and show the class of the sample using different colors (red for class 0 & blue for class 1)



6) (optional) Plot the boundary of the classifier.



7) Determine the accuracy of the learned method over the training data.

$$accuracy = 89\%$$

**F. (Optional) Weighted Linear Regression**

In part B, you analyzed the effect of outlier on linear regression method. In this part, you should apply weighted linear regression on the data file used in part B which includes the outliers.

1) What is the closed form solution for estimating the parameters in WLR?

$$\hat{\beta}_{WLS} = (\mathbf{x}^T \mathbf{W} \mathbf{x})^{-1} \mathbf{x}^T \mathbf{W} \mathbf{y}$$

2) Propose a weighting function to decrease the effect of outlier and explain the reasons that you select that function.

با استفاده از فرمول زیر به هر داده وزنی بین صفر و یک نسبت داده می شود:

$$w^{(i)} = \exp \left( -\frac{(x^{(i)} - x)^2}{2\tau^2} \right)$$

وزنی که به هر داده از دیتاست برای ایجاد مدل برای داده X انتساب می شود براساس نزدیکی هر داده از دیتاست به داده X می باشد، یک پنجره به مرکز X و پهنای  $\tau$  زده می شود و داده های نزدیکتر به X وزن بیشتر و داده های دورتر از X که در پنجره مدل قرار نمی گیرند وزن کمتری خواهند گرفت.

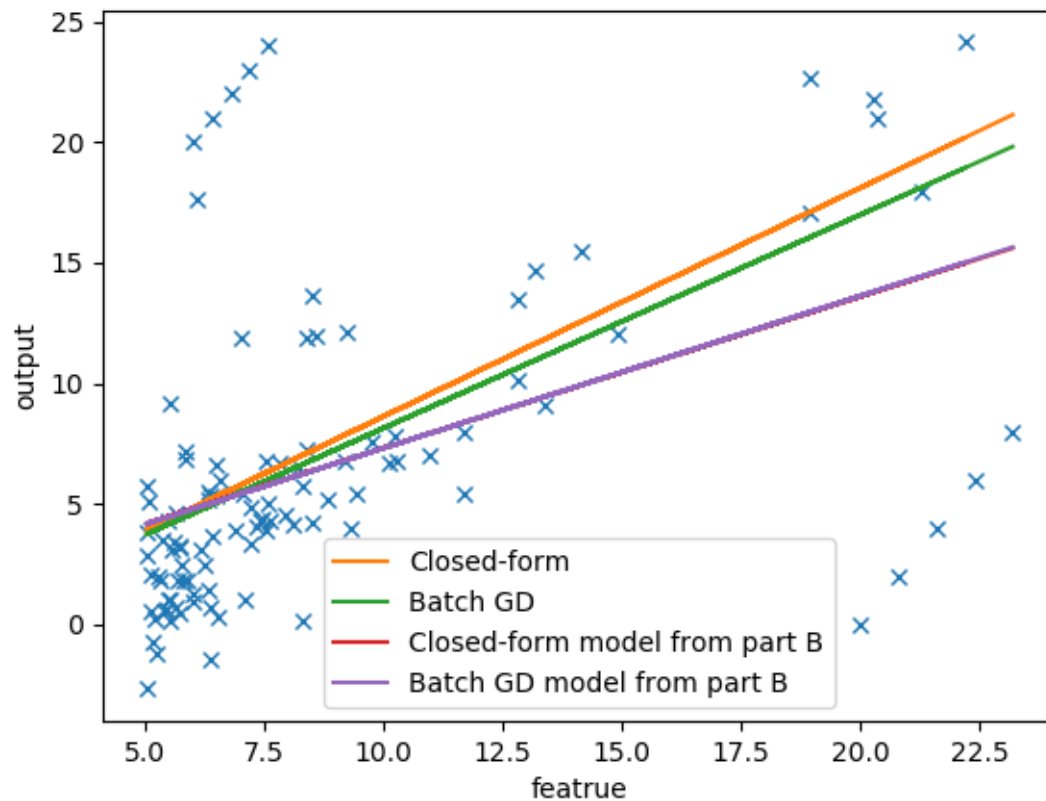
در این رابطه  $\tau$  یک free parameter است، برای مقادیر مختلف امتحان کردیم و  $\tau = 5$  انتخاب شد که به ازای پارامترهای تخمین زده شده، کمترین مقدار MSE بدست آمد.

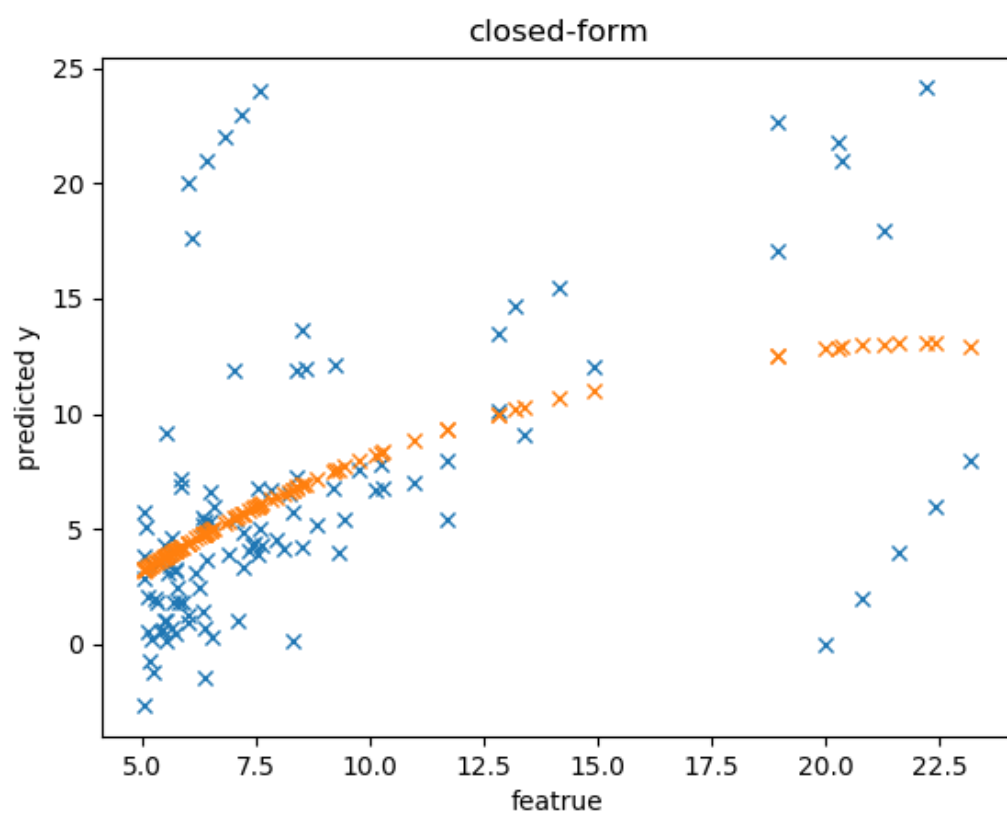
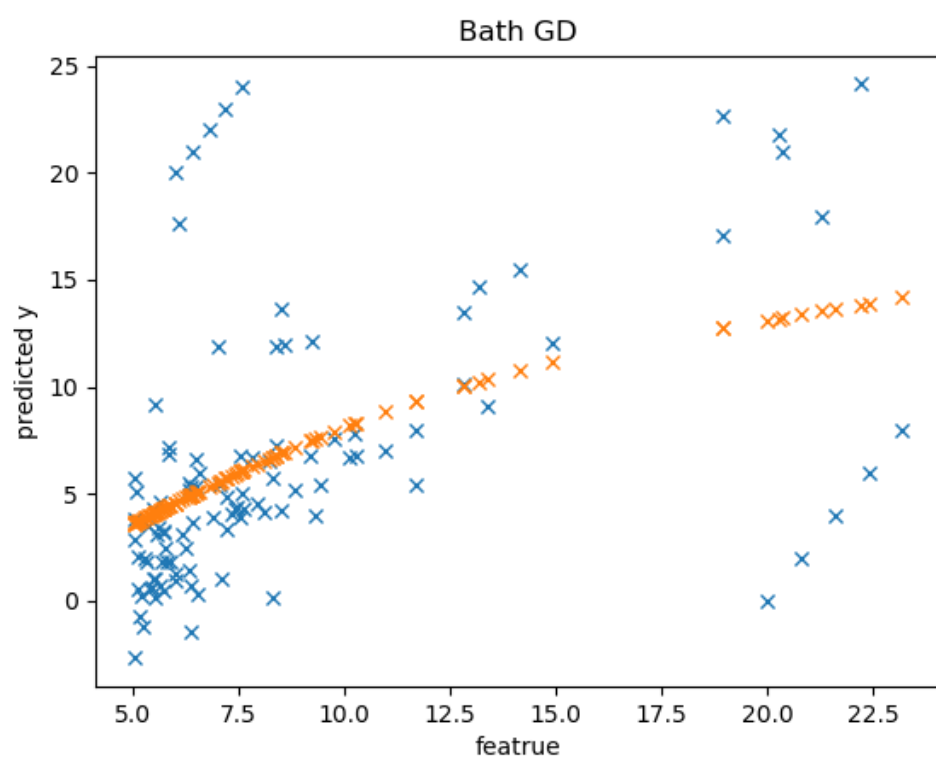
	MSE Closed-form	MSE Batch GD	Average MSE
$\tau = 0.5$	55.8708	15.7412	35.8060
$\tau = 1.0$	42.7329	16.1163	29.4246
$\tau = 4.5$	17.4461	16.4230	16.9345
$\tau = 4.6$	17.3557	16.4118	16.8837
$\tau = 4.7$	17.2701	16.3994	16.8347
$\tau = 4.8$	17.1884	16.3860	16.7872
$\tau = 4.9$	17.1102	16.3715	16.7408
$\tau = 5.0$	17.0350	16.3559	<b>16.6954</b>

- 3) Apply weighted linear regression using your suggested weighting function and:
- Closed form solution of weighted linear regression.
  - Batch Gradient descent (1500 iterations).

	$\theta_0$	$\theta_1$
<b>Closed-form</b>	-0.8594	0.9482
<b>Batch GD</b>	-0.7156	0.8851

- 4) Plot the data (including the outliers) and superimpose the fitted model in part B and the fitted models in this part onto the figures.





5) Compare the results in this part with part B.

با در نظر گرفتن وزن کمتر برای داده‌های باعث شد تا مدل های رسم شده نسبت به قسمت B کمتر به سمت داده‌های پرت متمایل شوند.

6) When it is better to use WLR method and when it is not?

بهتر است زمانی از WLR استفاده شود که وزن دقیق هر داده را بدانیم، زیرا با وزن دهی اشتباه داده‌ها ممکن است پارامترهای مدل به اشتباه تخمین زده شوند و دیگر مدل روی داده‌ها fit نباشد.