Justin Lam
77333383

# CIVL 498A HW-1: Batch GD, SGD, Regularization.

**Note**: For hand calculation problems, you are supposed to write down your answers on a piece of paper and then submit the scanned version of the paper(s). Show your work by writing out derivations if there is any. For coding problems, you can either use Google Colab or Jupyter Notebook (on your own machine) to write and run the code. When submitting, only submit the ipynb file, and make sure when "run all" is clicked, all required results from the questions are shown (this is how I will grade your answer. If when I "run all" and your code crashes, it will be deemed as wrong).

1. **(Hand calculation)** Assume a constant learning rate (alpha) of 1; use data in the following table:

| **Observations** | **$x_1$** | **$x_2$** | **$x_3$** | **$x_4$** | **y** |
|---|---|---|---|---|---|
| **1** | 2 | 2 | 5 | 2 | 10 |
| **2** | 1 | 5 | 10 | 5 | 20 |
| **3** | 5 | 2 | 10 | 5 | 10 |

1.1 Use stochastic gradient descent to find the hypothesis describing the relationship between y and the variables ($x_1$, $x_2$, $x_3$, $x_4$), plus another variable representing the intercept ($x_0$). <u>You should perform SGD for 1 epoch. Initial guesses of parameters (i.e., thetas) are all 0.</u>

look over entire data set

1.2 Now that you have a hypothesis, you are not satisfied with the performance and decide to improve the performance by using ridge regression. <u>Assume lambda to be 1, and perform batch GD for 1 epoch. You can disregard the theta values you calculated from 1.1 and start over. Initial guesses of parameters (i.e., thetas) are all 0.</u>

1.1

Guess: $\theta_0 = \theta_1 = \theta_2 = \theta_3 = \theta_4 = 0$

$x_0 = 1 \quad \alpha = 1$

Then: $\theta_j := \theta_j - \alpha \times x_j \cdot \underbrace{\left[ h_\theta(x)^{(i)} - y^{(i)} \right]}_{(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4) - (y)}$

1st example: (2, 2, 5, 2, 10)

$\theta_0 := 0 - 1[0 - 10] = 10$

$\theta_1 = 0 - 2(0 - 10) = 20$

$\theta_2 = 20$

$\theta_3 = 50$

$\theta_4 = 20$

2nd example: (1, 5, 10, 5, 20)

$\theta_0 := 10 - 1[10 + 20 \times 1 + 20 \times 5 + 50 \times 10 + 20 \times 5 - 20)$

$= 10 - 1[730 - 20] = 10 - 1(710) = -700$

$\theta_1 = 20 - 1 \times [710] \times 1 = -690$

$\theta_2 = 20 - 1 \times 710 \times 5 = -3530$

$\theta_3 = 50 - 1 \times 710 \times 10 = -7050$

$\theta_4 = 20 - 1 \times 710 \times 5 = -3530$

$3^{rd}$ example: $(5, 2, 10, 5, 10)$

$$h_\theta(x) = -700 - 690(5) - 3530(2) - 7050(10) - 3530(5)$$
$$= -99360$$

$$h_\theta(x) - y = -99370$$

$$\theta_0 := -700 - 1 \times -99370 \times 1 = 98,670$$

$$\theta_1 := -690 - 1 \times -99370 \times 5 = 496,160$$

$$\theta_2 = -3530 - 1 \times -99370 \times 2 = 195,210$$

$$\theta_3 = -7050 - 1 \times -99370 \times 10 = 986,650$$

$$\theta_4 = -3530 - 1 \times -99370 \times 5 = 493,320$$

$$\boxed{\text{Hypothesis}_{SGD}: \quad y = 98,670 + 496,160X_1 + 195,210X_2 + 986,650X_3 + 493,320X_4}$$

## 1.2

$\lambda = 1$, 1 epoch, $\theta_0 = \theta_1 = \theta_2 = \theta_3 = \theta_4 = 0$

$h_\theta = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3 + \theta_4 X_4$

$$\theta_j := \theta_j - \alpha \sum_{i=1}^{m} \left[ h_\theta(x^{(i)}) - y^{(i)} \right] \cdot x_j^{(i)} - \alpha \cdot \lambda \cdot \theta_j$$

$$\theta_0 = 0 - 1 \times \left[ (0-10)(1) + (0-20)(1) + (0-10)(1) \right] - 1 \times 1 \times 0$$

$$\theta_0 = -1 \times (-40) - 0 = 40$$

$$\theta_1 = 0 - 1 \left[ (0-10)(2) + (0-20)(1) + (0-10)(5) \right] - 1 \times 1 \times 0$$

$$\theta_1 = -1 \times (-90) - 0 = 90$$

$$\theta_2 = 0 - 1 \left[ (0-10)(2) + (0-20)(5) + (0-10)(2) \right] - 1 \times 1 \times 0$$

$$\theta_2 = -1 \times (-140) - 0 = 140$$

$$\theta_3 = 0 - 1 \left[ (0-10)(5) + (0-20)(10) + (0-10)(10) \right] - 1 \times 1 \times 0$$

$$\theta_3 = -1 \times (350) - 0 = 350$$

$$\theta_4 = 0 - 1 \left[ (0-10)(2) + (0-20)(5) + (0-10)(5) \right] - 1 \times 1 \times 0$$

$$\theta_4 = -1 \times (-170) - 0 = 170$$

$$\boxed{\text{Hypotheses Ridge: } y = 40 + 90 X_1 + 140 X_2 + 350 X_3 + 170 X_4}$$

2. **(Hand calculation)** You are the owner of an ironwork shop, and you have collected the productivity data of novice and experienced workers for various contracts you won. For experienced workers, you noticed their *productivity* has a linear relationship with respect to the *complexity of the work piece*, and the *number of days until deadline*. For novice workers, you noticed their *productivity* also has a linear relationship with respect to *complexity of the work piece*, and *number of days until deadline*.

You have productivity data of both experienced and novice workers collected in the following tables. For the questions below, assume a constant learning rate (alpha) of 0.01.

2.1. For experienced workers, use stochastic gradient descent to find the hypothesis describing the relationship between productivity and the complexity of work piece and number of days until deadline, plus another parameter representing the intercept. You should perform SGD for 2 epochs. Initial guess of parameter for complexity of work piece = 0, for number of days until deadline = -1, for the intercept = 10

| **Observations** | **Complexity of work piece** | **# of days until deadline** | **Productivity** |
|---|---|---|---|
| **1** | 2 | 6 | 2 |
| **2** | 1 | 5 | 7 |
| **3** | 5 | 2 | 1 |
| **4** | 2 | 3 | 8 |
| **5** | 4 | 4 | 0 |

create 2D List! (2,6) (1,5) (5,2) (2,3)

2.2. For novice workers, use batch gradient descent to find the hypothesis describing the relationship between productivity and the complexity of work piece and number of days until deadline, plus another parameter representing the intercept. You should perform batch GD for 2 epochs. Initial guess of parameter for complexity of work piece = 1, for number of days until deadline = -2, for intercept = 20

| **Observations** | **Complexity of work piece** | **# of days until deadline** | **Productivity** |
|---|---|---|---|
| **1** | 2 | 3 | 4 |
| **2** | 1 | 5 | 2.5 |
| **3** | 2 | 2 | 6 |
| **4** | 2 | 3 | 4 |
| **5** | 3 | 1 | 5.5 |

## 2.1 2 epochs

Guess: $\alpha = 0.01$

$\theta_1 = 0 \quad \theta_2 = -1 \quad \theta_0 = 10$

$$\theta_j := \theta_j - \alpha \times x_j \underbrace{\left[h_\theta(x)^{(i)} - y^{(i)}\right]}_{(\theta_0 + \theta_1 x_1 + \theta_2 x_2) - (y)}$$

1st example: $(2, 6, 2)$

$h_\theta(x) = (10 + 0 - 6) = 4$

$h_\theta(x) - y = 4 - 2 = 2$

$\theta_0 = 10 - 0.01 \times 1\,(2) = 9.98$

$\theta_1 = 0 - 0.01 \times 2\,(2) = -0.04$

$\theta_2 = -1 - 0.01 \times 6\,(2) = -1.12$

2nd ex: $(1, 5, 7)$

$h_\theta(x) - y = (9.98 - 0.04 \times 1 - 1.12 \times 5) - 7 = -2.66$

$\theta_0 = 9.98 - 0.01 \times 1 \times (-2.66) = 10.01$

$\theta_1 = -0.04 - 0.01 \times 1 \times (-2.66) = -0.0134$

$\theta_2 = -1.12 - 0.01 \times 5\,(-2.66) = -0.987$

$3^{rd}$ ex: $(5,2,1)$

$$h_0(x) - y = (10.01 - 0.0134 \times 5 - 0.987 \times 2) - 1 = 6.97$$

$$\theta_0 = 10.01 - 0.01\,(1)(6.97) = 9.94$$

$$\theta_1 = -0.0134 - 0.01\,(5)(6.97) = -0.362$$

$$\theta_2 = -0.987 - 0.01\,(2)(6.97) = -1.126$$

$4^{th}$ ex: $(2,3,8)$

$$h_0(x) - y = (9.94 - 0.362 \times 2 - 1.126 \times 3) - 8 = -2.162$$

$$\theta_0 = 9.94 - 0.01\,(1)(-2.162) = 9.961$$

$$\theta_1 = -0.362 - 0.01\,(2)(-2.162) = -0.319$$

$$\theta_2 = -1.126 - 0.01\,(3)(-2.162) = -1.06$$

$5^{th}$ ex: $(4,4,0)$

$$h(x) - y = (9.961 - 0.319 \times 4 - 1.06 \times 4) - 0 = 4.445$$

$$\theta_0 = 9.961 - 0.01 \times 1\,(4.445) = 9.916$$

$$\theta_1 = -0.319 - 0.01 \times 4\,(4.445) = -0.496$$

$$\theta_2 = -1.06 - 0.01 \times 4\,(4.445) = -1.238$$

2nd epoch ... ~

↓

$6^{th}$ ex: $(2,6,2)$

$h(x)-y = (9.916 - 0.496 \times 2 - 1.238 \times 6) - 2 = -0.504$

$\theta_0 = 9.916 - 0.01 \times 1 \times -0.504 = 9.92$

$\theta_1 = -0.496 - 0.01 \times 2 \times -0.504 = -0.486$

$\theta_2 = -1.238 - 0.01 \times 6 \times -0.504 = -1.208$

$7^{th}$ ex: $(1,5,7)$

$h_\theta(x)-y = (9.92 - 0.486 \times 1 - 1.208 \times 5) - 7 = -3.606$

$\theta_0 = 9.92 - 0.01 \times 1 \times -3.606 = 9.956$

$\theta_1 = -0.486 - 0.01 \times 1 \times -3.606 = -0.449$

$\theta_2 = -1.208 - 0.01 \times 5 \times -3.606 = -1.028$

$8^{th}$ ex: $(5,2,1)$

$h(x)-y = (9.956 - 0.449 \times 5 - 1.028 \times 2) - 1 = 4.655$

$\theta_0 = 9.956 - 0.01 \times 4.655 \times 1 = 9.909$

$\theta_1 = -0.449 - 0.01 \times 4.655 \times 5 = -0.682$

$\theta_2 = -1.028 - 0.01 \times 4.655 \times 2 = -1.121$

$9^{th}$ ex: (2,3,8)

$$h(x)-y = (9.909-0.682\times 2-1.121\times 3)-8 = -2.818$$

$$\theta_0 = 9.909-0.01\times -2.818\times 1 = 9.937$$

$$\theta_1 = -0.682-0.01\times -2.818\times 2 = -0.626$$

$$\theta_2 = -1.121-0.01\times -2.818\times 3 = -1.036$$

$10^{th}$ ex: (4,4,0)

$$h(x)-y = (9.937-0.626\times 4-1.036\times 4)-0 = 3.289$$

$$\theta_0 = 9.937-0.01\times 3.289\times 1 = 9.904$$

$$\theta_1 = -0.626-0.01\times 3.289\times 4 = -0.757$$

$$\theta_2 = -1.036-0.01\times 3.289\times 4 = -1.168$$

Final Hypothesis SGD:

$$y = 9.904-0.757X_1-1.168X_2$$

where $y$ = productivity

$X_1$ = complexity of work piece

$X_2$ = # of days until deadline

## 2.2 2 epochs

Guess: $\alpha = 0.01$

$\theta_0 = 20 \quad \theta_1 = 1 \quad \theta_2 = -2$

$$\theta_j := \theta_j - \alpha \sum_{i=1}^{M} \left[ h_\theta(x)^{(i)} - y^{(i)} \right] \cdot x_j^{(i)}$$

$$\theta_j := \theta_j - \alpha \sum_{i=1}^{M} \left[ \left( \theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)} \right) - y^{(i)} \right] x_j^{(i)}$$

1st epoch:

$$\theta_j = \theta_j - \alpha \sum_{i=1}^{m} \left( 20 + x_1^{(i)} - 2x_2^{(i)} - y^{(i)} \right) \left( x_j^{(i)} \right)$$

$$\theta_0 = 20 - 0.01 \big[ (20+2-2\times3-4)(1) + (20+1-2\times5-2.5)(1) + (20+2-2\times2-6)(1) + (20+2-2\times3-4)(1) + (20+3-2\times1-5.5)(1) \big]$$

$$\theta_0 = 20 - 0.01(60) = 19.4$$

$$\theta_1 = 1 - 0.01 \big[ (20+2-2\times3-4)(2) + (20+1-2\times5-2.5)(1) + (20+2-2\times2-6)(2) + (20+2-2\times3-4)(2) + (20+3-2\times1-5.5)(3) \big]$$

$$\theta_1 = 1 - 0.01(127) = -0.27$$

$$\theta_2 = -2 - 0.01\big[(20+2-2\times3-4)(3) + (20+1-2\times5-2.5)(5)$$
$$+ (20+2-2\times2-6)(2) + (20+2-2\times3-4)(3)$$
$$+ (20+3-2\times1-5.5)(1)\big]$$

$$\theta_2 = -2 - 0.01(154) = -3.54$$

2nd epoch:

$$\theta_j := \theta_j - \alpha \sum_{i=1}^{m} \left[\left(\theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)}\right) - y^{(i)}\right] x_j^{(i)}$$

$$\theta_j := \theta_j - 0.01 \sum_{i=1}^{m} \left[(19.4 - 0.27x_1 - 3.54x_2) - y\right] x_2^{(i)}$$

$$\theta_0 := 19.4 - 0.01\big[(19.4-0.27\times2-3.54\times3-4)(1)$$
$$+$$
$$(19.4-0.27\times1-3.54\times5-2.5)(1)$$
$$+$$
$$(19.4-0.27\times2-3.54\times2-6)(1)$$
$$+$$
$$(19.4-0.27\times2-3.54\times3-4)(1)$$
$$+$$
$$(19.4-0.27\times3-3.54\times1-5.5)(1)\big]$$

$$\theta_0 = 19.4 - 0.01(22.74) = 19.173$$

$$\theta_1 := -0.27 - 0.01\Big[(19.4 - 0.27\times 2 - 3.54\times 3 - 4)(2)$$
$$+ (19.4 - 0.27\times 1 - 3.54\times 5 - 2.5)(1)$$
$$+ (19.4 - 0.27\times 2 - 3.54\times 2 - 6)(2)$$
$$+ (19.4 - 0.27\times 2 - 3.54\times 3 - 4)(2)$$
$$+ (19.4 - 0.27\times 3 - 3.54\times 1 - 5.5)(3)\Big]$$

$$\theta_1 = -0.27 - 0.01(56.1) = -0.831$$

$$\theta_2 := -3.54 - 0.01\Big[(19.4 - 0.27\times 2 - 3.54\times 3 - 4)(3)$$
$$+ (19.4 - 0.27\times 1 - 3.54\times 5 - 2.5)(5)$$
$$+ (19.4 - 0.27\times 2 - 3.54\times 2 - 6)(2)$$
$$+ (19.4 - 0.27\times 2 - 3.54\times 3 - 4)(3)$$
$$+ (19.4 - 0.27\times 3 - 3.54\times 1 - 5.5)(1)\Big]$$

$$\theta_2 := -3.54 - 0.01(41.2) = -3.952$$

$$\boxed{\begin{array}{l}\text{Final Hypotesis Batch:}\\ y = 19.173 - 0.831X_1 - 3.952X_2\\ \text{where } y \text{ is productivity, } X_1 \text{ is complexity of work, } X_2 \text{ is days until deadline}\end{array}}$$

2.3. You won another contract with the complexity of the work piece as 3, and there is still 5 days left until the deadline. Using the hypothesis you trained from previous 2 questions, what is the productivity of novice workers, and what is the productivity of experienced workers?

$\boxed{\text{Novice workers : prod} = -3.08}$

$$y = 19.173 - 0.831 \times 3 - 3.952 \times 5 = -3.08$$

$\boxed{\text{Experienced workers : prod} = 1.793}$

$$y = 9.904 - 0.757 \times 3 - 1.168 \times 5 = 1.793$$

3. **(Coding problem)** Rework question 2 in Google Colab or using Jupyter Notebook on your own machine for the following questions:

3.1. Redo question 2.1, using the same data and assumptions; however, perform SGD for 10, 100, and 1000 epochs and report their resulted hypothesis, in the form of printing out values of thetas. For this question, you are not supposed to use any Python packages.

3.2. Redo question 2.2, using the same data and assumptions; however, perform batch GD for 10, 100, and 1000 epochs and report their resulted hypothesis, in the form of printing out values of thetas. For this question, you are not supposed to use any Python packages.

3.3. Redo question 2.1 using the sklearn.linear_model.LinearRegression class with all default parameters; redo question 2.2 using the sklearn.linear_model.Ridge class, with all default parameters. You don't have to consider the number of epochs trained, the Scikit Learn classes will take care of this for you. Report both results in the form of printing out values of thetas.

10,000 epochs