David Kwak
kwakd@oregonstate.edu

HW2 - Linear Models for Regression and Classification

CS 434 HW 2

Q1. Show that the MLE for this model minimizes the sum of the absolute errors (SAE):

$$SAE(w) = \sum_{i=1}^{N} |y_i - w^T x_i|$$

Laplace Dist:

- $y_i \sim \text{Laplace}(\mu = w^T x_i, b) \rightarrow p(y_i | x_i, w) = \frac{1}{2b} e^{-\frac{|y_i - w^T x_i|}{b}}$

↳ Likelihood: $(\mu = w^T x_i, b)$

$$P(y|x,w) = \prod_{i=1}^{n} P[y_i | x_i, w]$$

$$= \prod_{i=1}^{n} \left(\frac{1}{2b} e^{-\frac{|y_i - w^T x_i|}{b}} \right)$$

$$= \frac{1}{(2b)^n} \cdot e^{-\frac{1}{b} \sum_{i=1}^{n} |y_i - w^T x_i|}$$

↳ When $\sum_{i=1}^{N} |y_i - w^T x_i|$ increases

↳ Then $-\frac{1}{b} \sum_{i=1}^{n} |y_i - w^T x_i|$ decreases

↳ For example: $e^{-\frac{1}{b} \sum_{i=1}^{n} |y_i - w^T x_i|}$ will decrease

↳ Therefore the likelihood is proportional to $\sum_{i=1}^{n} [y_i - w^T x_i]$.

- If b is fixed, then the likelihood is maximized when $[-\sum_{i=1}^{n} (y_i - w^T x_i]$ is also maximized.

↳ For example: $\sum_{i=1}^{n} |y_i - w^T x_i|$ is maximized

↳ Therefore the MLE of w also maximize $\sum_{i=1}^{n} |y_i - w^T x_i|$

Q2.

- Recall = $\frac{\#\text{ True Positive}}{\#\text{ True Positive} + \#\text{ False Negatives}}$

- Precision = $\frac{\#\text{ True Positive}}{\#\text{ True Positives} + \#\text{ False Positives}}$

y	p(y|x)	t=0	t=0.2	t=0.4	t=0.6	t=0.8	t=1
0	0.1	1	0	0	0	0	0
0	0.1	1	0	0	0	0	0
0	0.25	1	1	0	0	0	0
1	0.25	1	1	0	0	0	0
0	0.3	1	1	0	0	0	0
0	0.33	1	1	0	0	0	0
1	0.4	1	1	0	0	0	0
0	0.52	1	1	1	0	0	0
0	0.53	1	1	1	0	0	0
1	0.7	1	1	1	1	0	0
1	0.8	1	1	1	1	0	0
0	0.85	1	1	1	1	1	0
1	0.9	1	1	1	1	1	0
1	0.9	1	1	1	1	1	0
1	0.95	1	1	1	1	1	0
1	1.0	1	1	1	1	1	0

Let : TN = True Negative FN = False negative TP = True Pos. FP = False pos.

- t=0

	Predicted No (0)	Predicted Yes (1)
Actual No (0)	TN 0	FP 8
Actual Yes (1)	FN 0	TP 8

↳ Recall = $\frac{TP}{TP+FN} = \frac{8}{8+0} = 1$

↳ Precision = $\frac{TP}{TP+FP} = \frac{8}{8+8} = \frac{8}{16} = \frac{1}{2} = 0.5$

- $t = 0.2$

	Pred 0	Pred 1
Act 0	TN 2	FP 6
Act 1	FN 0	TP 8

↳ Recall $= \frac{8}{8+0} = 1$

↳ Prec. $= \frac{8}{8+6} = \frac{84}{147} = \frac{4}{7} = 0.571$

- $t = 0.4$

	Pred 0	Pred 1
Act 0	TN 5	FP 3
Act 1	FN 2	TP 6

↳ Recall $= \frac{6}{6+2} = \frac{6}{8} = \frac{3}{4} = 0.75$

↳ Prec $= \frac{6}{6+3} = \frac{6}{9} = \frac{2}{3} = 0.667$

- $t = 0.6$

	Pred 0	Pred 1
Act 0	TN 7	FP 1
Act 1	FN 2	TP 6

↳ Recall $= \frac{6}{6+2} = \frac{6}{8} = \frac{3}{4} = 0.75$

↳ Prec. $= \frac{6}{6+1} = \frac{6}{7} = 0.857$

- $t = 0.8$

	Pred 0	Pred 1
Act 0	TN 7	FP 1
Act 1	FN 4	TP 4

↳ Recall $= \frac{4}{4+4} = \frac{4}{8} = \frac{1}{2} = 0.5$

↳ Prec $= \frac{4}{4+1} = \frac{4}{5} = 0.8$

- $t = 1$

	Pred 0	Pred 1
Act 0	TN 8	FP 0
Act 1	FN 8	TP 0

↳ Recall $= \frac{0}{0+8} = \frac{0}{8} = 0$

↳ Prec $= \frac{0}{0+0} = \frac{0}{0} =$ undef.

Q4.

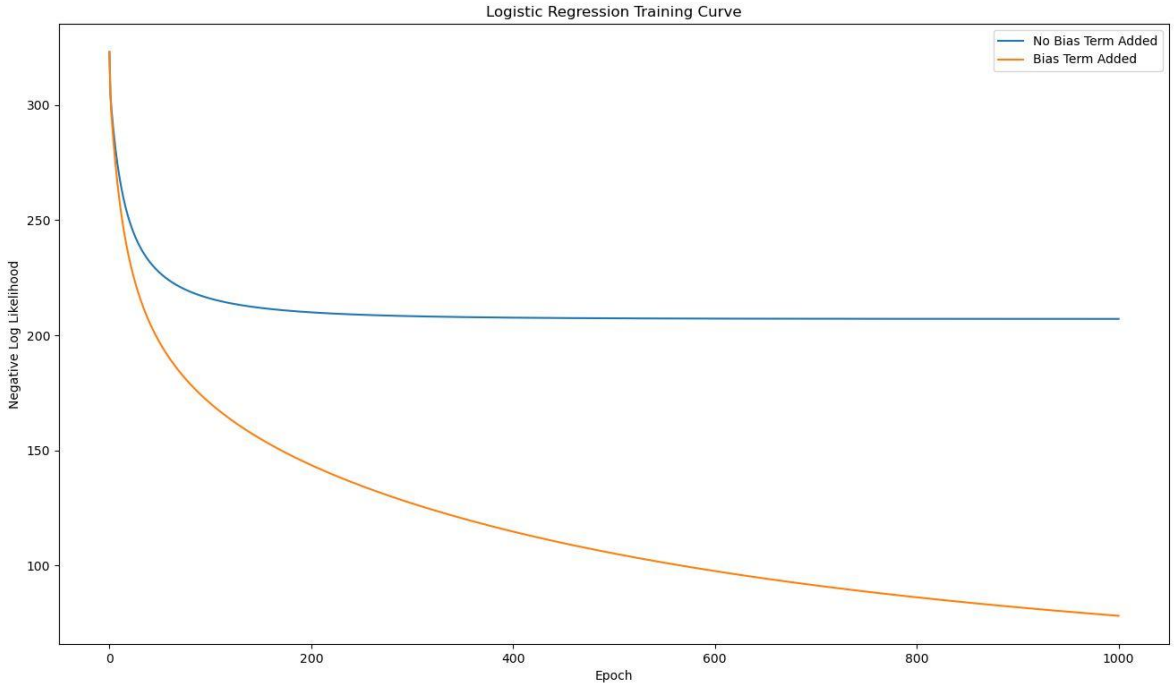
- Weight Vector: [-0.2464, 0.8677, 0.2008, 0.2785, -0.676]
- Accuracy: 86.27%

Q5.

- Weight Vector: [-3.4144, 0.08, 0.4192, 0.2177, 0.2745, -0.0528]
- Accuracy: 96.35%
- It did make a meaningful difference in the terms of how the accuracy went up by around 10%.

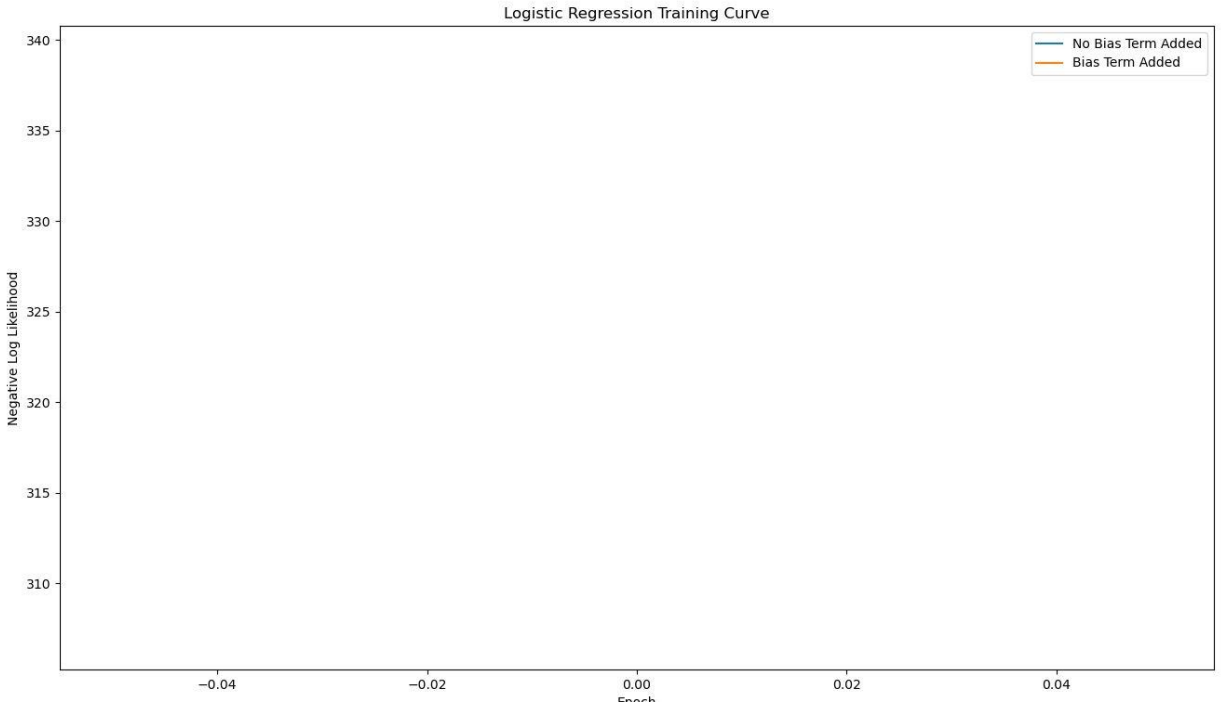
Q6.

- Based on the graph for 0.0001 the base step_size it looks as if the graph would continue to drop if the max_iteration was set higher when the basic term is added.
- **0.0001**



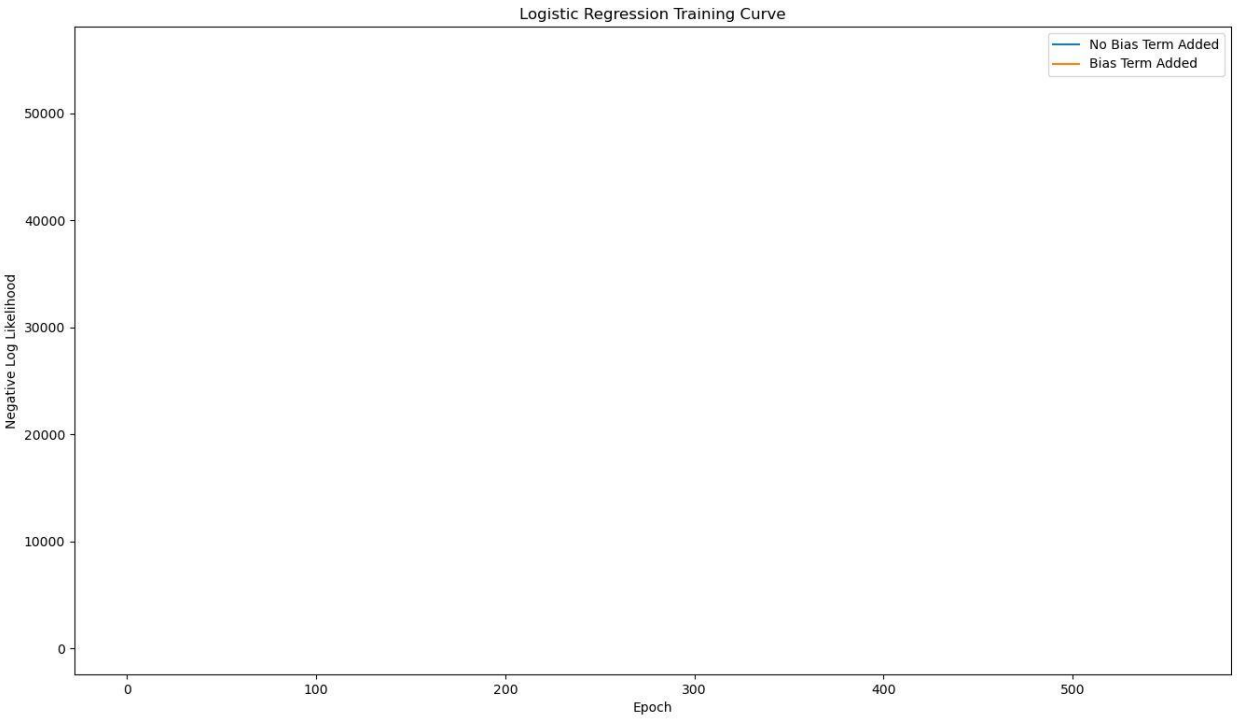
-
- Training Accuracy (NO BIAS): 86.27%
- Training Accuracy (BIAS): 96.35%

-
- **1**



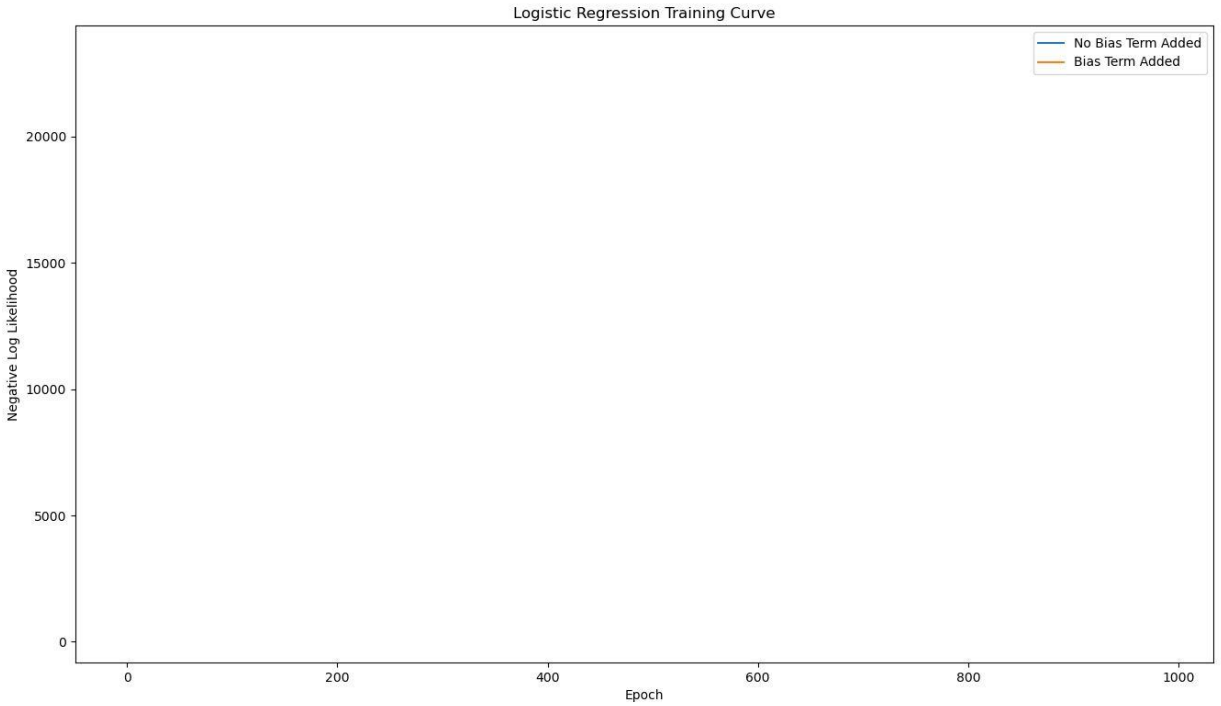
-
- Training Accuracy (NO BIAS): 82.4%
- Training Accuracy (BIAS): 87.98%

- **0.1**



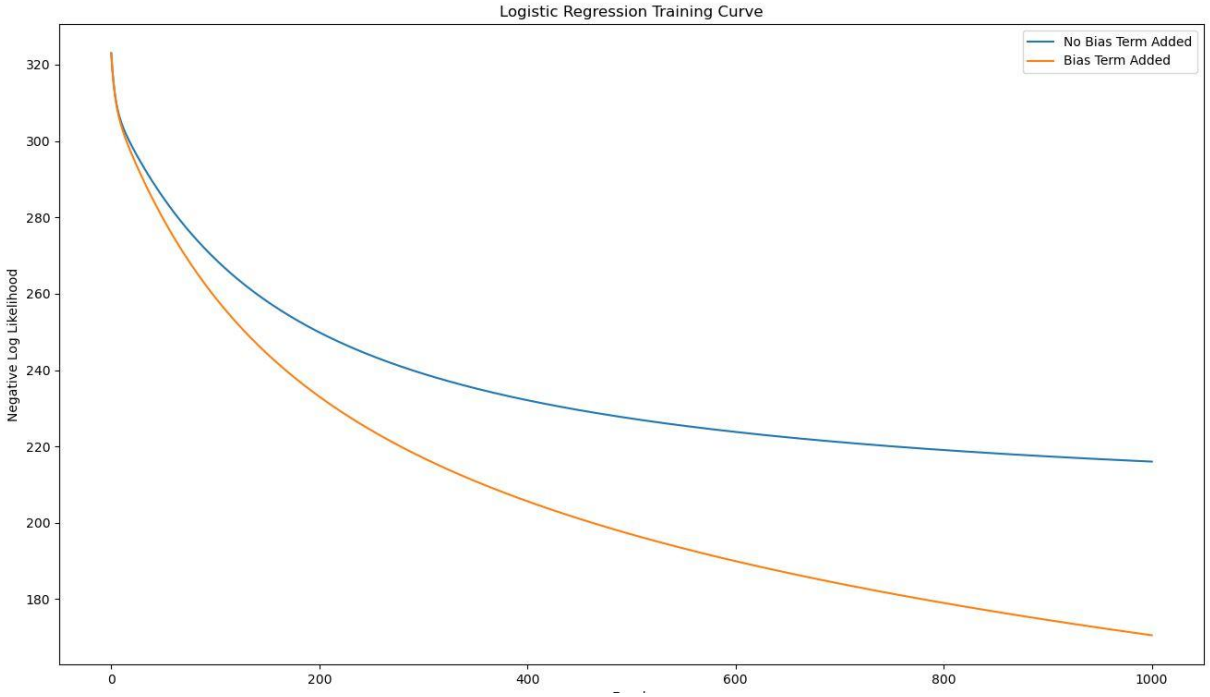
  -
  - Training Accuracy (NO BIAS): 75.54%
  - Training Accuracy (BIAS): 95.92%
-

- **0.01**



-
- Training Accuracy (NO BIAS): 84.12%
- Training Accuracy (BIAS): 97.0%

- **0.00001**



  - Training Accuracy (NO BIAS): 85.62%
  - Training Accuracy (BIAS): 90.77%
- The trends whenever the value was above 0.0001 I'd get a "runtimewarning" however as the number got less you can see the bias term's model gets lower and lower.

Q7.

- At first my performance was very poor on the kaggle, so it wasn't matching my performance with these means and standard deviations. However after changing my w value to the w from X_train the means and standard deviations were matching to the kaggle submission.

Debriefing

1. Approximately how many hours did you spend on this assignment?

- Approximately around 20 hours was spent on this assignment.

2. Would you rate it as easy, moderate, or difficult?

- I would rate this as somewhat moderate near difficult just because there are still parts of the topic I don't understand in terms of coding, however with some help I was able to understand more and more.

3. Did you work on it mostly alone or did you discuss the problems with others?

- Same as the last homework I used many resources from the canvas along with asking a peer for some ideas and online resources.

4. How deeply do you feel you understand the material it covers (0%–100%)?

- I wanna say around 65% because there are still some areas of the material I don’t understand fully, however I’m going to review them before the midterm.

5. Any other comments?

- N/A