Homework 2 Gretel Rajamoney

Q1:

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

$Y_i \sim \text{Laplace}(\mu = \omega^T x_i, b)$

$$P(y_i | x_i, \omega) = \frac{1}{2b} \cdot e^{-\frac{|y_i - \omega^T x_i|}{b}}$$

therefore, the likelihood

$$P(y|X, \omega) = \prod_{i=1}^{n} P(y_i | x_i, \omega)$$

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

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

therefore, as $\sum_{i=1}^{n} |y_i - \omega^T x_i|$ is increasing,

as $-1/b \cdot \sum_{i=1}^{n} |y_i - \omega^T x_i|$ is decreasing,

as $e^{-\frac{1}{b} \cdot \sum_{i=1}^{n} |y_i - \omega^T x_i|}$ is decreasing

therefore, likelihood is proportional to $\sum_{i=1}^{n} |y_i - \omega^T x_i|$

if b is fixed, likelihood is maximized when $\left(-\sum_{i=1}^{n} |y_i - \omega^T x_i|\right)$ is maximized, or when $\left(\sum_{i=1}^{n} |y_i - \omega^T x_i|\right)$ is minimized

in conclusion, the MLE of ω also minimizes $\sum_{i=1}^{n} |y_i - \omega^T x_i|$

Scanned with CamScanner

Q2:

		t = 0	t = 0.2	t = 0.4	t = 0.6	t = 0.8	t = 1
y(actual)	P(y|x)	←		y(predicted)			→
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	0.55	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	1	1	1	1	1	0

t = 0	t = 0.2	t = 0.4
TP = 8, TN = 0, FP = 8, FN = 0	TP = 8, TN = 2, FP = 6, FN = 0	TP = 6, TN = 5, FP = 3, FN = 2
Recall = 8/(8+0) = 1	Recall = 8/(8+0) = 1	Recall = 6/(6+2) = 0.75
Precision = 8/(8+8) = 0.5	Precision = 8/(8+6) = 4/7	Precision = 6/(6+3) = 2/3

t = 0.6	t = 0.8	t = 1
TP = 6, TN = 7, FP = 1, FN = 2	TP = 4, TN = 7, FP = 1, FN = 4	TP = 0, TN = 8, FP = 0, FN = 8
Recall = 6/(6+2) = 0.75	Recall = 4/(4+4) = 0.5	Recall = 0/(0+8) = 0
Precision = 6/(6+1) = 6/7	Precision = 4/(4+1) = 4/5	Precision = 0/(0+0) = $\varnothing$

Scanned with CamScanner

Q4:

weight vector: [-0.2464, 0.8677, 0.2008, 0.2785, -0.6761,
-0.3325, 0.4337, -0.3499]

train accuracy: 86.27%

Q5:

weight vector: [0.3988, 0.178, 0.342, 0.433, -0.0306,
0.3404, 0.2459, 0.1911, -0.9197, -0.9197,
-0.9197, -0.9197, -0.9197, -0.9197,
-0.9197, -0.9197]

train accuracy: 97.21%

The accuracy when including the bias, is significantly higher than the accuracy without the bias.

Scanned with CamScanner

Layout References Mailings Review View Help

Paragraph Spacing

Effects

Set as Default

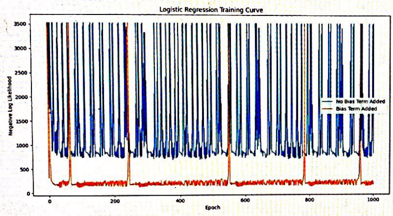
Colors Fonts

Document Formatting

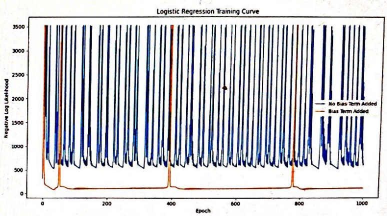
is temporarily stored on your computer. Save

Q6:

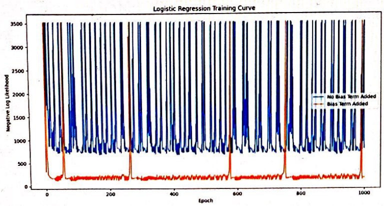
step size = 1



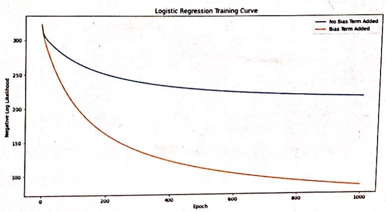
step size = 0.01



step size = 0.1



step size = 0.00001



Display Sett

Scanned with CamScanner

Q6:
As you can see on the plots, the bias and no bias lines diverge as the step size gets smaller. step sizes 0.0001 and 0.00001 were perfect for visualizing the divergence.

Q7:
Cross-validation was not a very ideal estimate of the programs actual performance on the leaderboard. The range of the means for the various k values of (2,3,4,5,10,20,50) is between 96.13% and 96.67%. Meanwhile my Kaggle submission recieved an accuracy of 91.379%. The trend that I did observe was that as k increases, the standard deviation increases as well.

Q8:
For my final submission, I biased my X_test variable utilizing the dummyArgument() function. I also inputted my biased train X variable (X_train-bias) within my trainLogistic() function. I also modified my step-size variable from 0.0001 to 0.00001, and my max_iters variable from 1000 to 10000.

Scanned with CamScanner

Debriefing

1. I spent approximately 15 hours on this assignment, the majority of those hours were for debugging.

2. Moderate in comparison to homework one.

3. Alone, heavily utilized the lecture slides pertaining to negative log likelihood, gradients, and regression.

4. 93%

5. none

Scanned with CamScanner