Sunghyun Kim
5362318
kimx5261@umn.edu

Homework 2
CSci 5525 Spring 23

2023-02-16

1 Derive Gradient Descent of $L(y_i|\mathbf{x}_i, \mathbf{w}) = y_i \log(\sigma(\mathbf{w}^T\mathbf{x}_i)) + (1-y_i)\log(\sigma(-\mathbf{w}^T\mathbf{x}_i))$

$$
\begin{aligned}
\frac{\partial L}{\partial w_j} &= \frac{\partial}{\partial w_j} y_i \log(\sigma(\mathbf{w}^T\mathbf{x}_i)) + \frac{\partial}{\partial w_j}(1-y_i)\log(\sigma(1-\mathbf{w}^T\mathbf{x}_i)) \\
&= \left(\frac{y_i}{\sigma(\mathbf{w}^T\mathbf{x}_i)} - \frac{1-y_i}{1-\sigma(\mathbf{w}^T\mathbf{x}_i)}\right)\frac{\partial}{\partial w_j}\sigma(\mathbf{w}^T\mathbf{x}_i) \\
&= \left(\frac{y_i}{\sigma(\mathbf{w}^T\mathbf{x}_i)} - \frac{1-y_i}{1-\sigma(\mathbf{w}^T\mathbf{x}_i)}\right)\sigma(\mathbf{w}^T\mathbf{x}_i)(1-\sigma(\mathbf{w}^T\mathbf{x}_i))x_j \text{ Algebraic manipulation} \\
&= \left(\frac{y-\sigma(\mathbf{w}^T\mathbf{x}_i)}{\sigma(\mathbf{w}^T\mathbf{x}_i)(1-\sigma(\mathbf{w}^T\mathbf{x}_i))}\right)\sigma(\mathbf{w}^T\mathbf{x}_i)(1-\sigma(\mathbf{w}^T\mathbf{x}_i))x_j \\
&= (y-\sigma(\mathbf{w}^T\mathbf{x}_i))x_j
\end{aligned}
$$

2 Logistic Regression with Full Gradient Descent

For both Logistic Regression & the most optimal η value was 0.00001.

Logistic Regression on Test Data & Error Rate by Fold											
1	2	3	4	5	6	7	8	9	10	Mean	SD
0.025	0.045	0.02	0.055	0.065	0.05	0.035	0.015	0.05	0.05	0.0375	0.0407

Please consult the code's output for more detailed result.

3 Derive Gradient Descent of $f(\mathbf{w}) = \frac{1}{2}||\mathbf{w}||_2^2 + C\sum_{i=1}^{n}\max(0, 1-y_i(\mathbf{w}^T\mathbf{x}_i + b))$

When $1 - y_i(\mathbf{w}^T\mathbf{x}_i + b) = 0$

$$
\begin{aligned}
\frac{\partial f(\mathbf{w})}{\partial w} &= \frac{\partial}{\partial w}\frac{1}{2}||\mathbf{w}||_2^2 \\
&= \mathbf{w}
\end{aligned}
$$

Otherwise,

$$
\begin{aligned}
\frac{\partial f(\mathbf{w})}{\partial w} &= \frac{\partial}{\partial w}\frac{1}{2}||\mathbf{w}||_2^2 + \frac{\partial}{\partial w}C\sum_{i=1}^{n} 1 - y_i(\mathbf{w}^T\mathbf{x}_i + b) \\
&= \mathbf{w} - C\sum_{i=1}^{n}(y_i x_i)
\end{aligned}
$$

4 Support Vector Machine with Full Gradient Descent

According to the iteration, most η for SVM seems to be between 0.00001, with C value of 1. Larger step sizes failed cost function to converge, which caused runtime warning.

Support Vector Machines on Test Data & Error Rate by Fold											
1	2	3	4	5	6	7	8	9	10	Mean	SD
0.0	0.0	0.025	0.0	0.0	0.025	0.0	0.0	0.025	0.0	0.0075	0.0115

Please consult the code's output for more detailed result.