Homework 2 - Practical Report

Reza Bayat and Arian Khorasani

November 2021

1 Question 1

$$\frac{C}{2}\sum_{j'=1}^{m}\left\|\mathbf{w}^{j'}\right\|_2^2$$

$$\frac{C}{2}\sum_{j'=1}^{m}\left\|\mathbf{w}^{j'}\right\|_2^2 = \frac{C}{2}\left((w_1{}^{j'})^2 + (w_2{}^{j'})^2 + \cdots + (w_m{}^{j'})^2\right)$$

Take the deviate of the regularization term with respect to $w_k{}^{j'}$.

$$\begin{aligned}
\frac{\partial}{\partial w_k{}^{j'}}\left(\frac{C}{2}\sum_{j'=1}^{m}\left\|\mathbf{w}^{j'}\right\|_2^2\right) &= \frac{C}{2}\frac{\partial}{\partial w_k{}^{j'}}\sum_{j'=1}^{m}\left\|\mathbf{w}^{j'}\right\|_2^2 \\
&= \frac{C}{2}\frac{\partial}{\partial w_k{}^{j'}}\left((w_1{}^{j'})^2 + (w_2{}^{j'})^2 + \cdots + (w_k{}^{j'})^2 + \cdots + (w_m{}^{j'})^2\right) \\
&= C w_k{}^{j'}
\end{aligned}$$

2 Question 2

This is the hing loss equation:

$$\frac{1}{n}\sum_{(\mathbf{x}_i,y_i)\in S}\sum_{j'=1}^{m}\mathcal{L}\left(\mathbf{w}^{j'};(\mathbf{x}_i,y_i)\right)$$

and

$$\mathcal{L}\left(\mathbf{w}^{j'};(\mathbf{x}_i,y_i)\right)=\left(\max\left\{0,2-\left(\left\langle\mathbf{w}^{j'},\mathbf{x}_i\right\rangle\right)\mathbb{1}\left\{y_i=j'\right\}\right\}\right)^2$$

We take the deviate of a loss term with respect to ${w_k}^{j'}$.

$$\begin{aligned}
\frac{\partial}{\partial {w_k}^{j'}}\mathcal{L}\left(\mathbf{w}^{j'};(\mathbf{x}_i,y_i)\right) &= \frac{\partial}{\partial {w_k}^{j'}}\left(\max\left\{0,2-\left(\left\langle\mathbf{w}^{j'},\mathbf{x}_i\right\rangle\right)\mathbb{1}\left\{y_i=j'\right\}\right\}\right)^2\\
&= 2\times\frac{\partial}{\partial {w_k}^{j'}}\left(\max\left\{0,2-\left(\left\langle\mathbf{w}^{j'},\mathbf{x}_i\right\rangle\right)\mathbb{1}\left\{y_i=j'\right\}\right\}\right)\times\max\left\{0,2-\left(\left\langle\mathbf{w}^{j'},\mathbf{x}_i\right\rangle\right)\mathbb{1}\left\{y_i=j'\right\}\right\}\\
&= 2\times\mathbf{x}_{i,k}\times\max\left\{0,2-\left(\left\langle\mathbf{w}^{j'},\mathbf{x}_i\right\rangle\right)\mathbb{1}\left\{y_i=j'\right\}\right\}
\end{aligned}$$

Take the derivative of the hinge loss term with respect to ${w_k}^{j'}$. We use above derivation in the following section.

$$\begin{aligned}
\frac{\partial}{\partial {w_k}^{j'}}\left(\frac{1}{n}\sum_{(\mathbf{x}_i,y_i)\in S}\sum_{j'=1}^{m}\mathcal{L}\left(\mathbf{w}^{j'};(\mathbf{x}_i,y_i)\right)\right) &= \frac{1}{n}\frac{\partial}{\partial {w_k}^{j'}}\left(\sum_{(\mathbf{x}_i,y_i)\in S}\sum_{j'=1}^{m}\mathcal{L}\left(\mathbf{w}^{j'};(\mathbf{x}_i,y_i)\right)\right)\\
&= \frac{2}{n}\times\mathbf{x}_{i,k}\times\max\left\{0,2-\left(\left\langle\mathbf{w}^{j'},\mathbf{x}_i\right\rangle\right)\mathbb{1}\left\{y_i=j'\right\}\right\}
\end{aligned}$$

3 Question 3

The implementation of the SVM is submitted in the GradeScope.

4 Question 4

Here are the four plots: one for each of Training Loss (figure 1), Training Accuracy (figure 2), Test Loss (figure 3), and Test Accuracy (figure 4). Each plot contains 3 curves, one for each value of C.

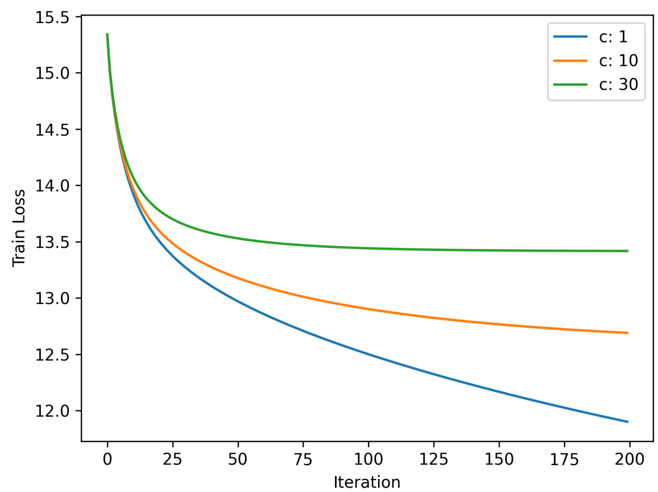

Figure 1: Training Loss Plot

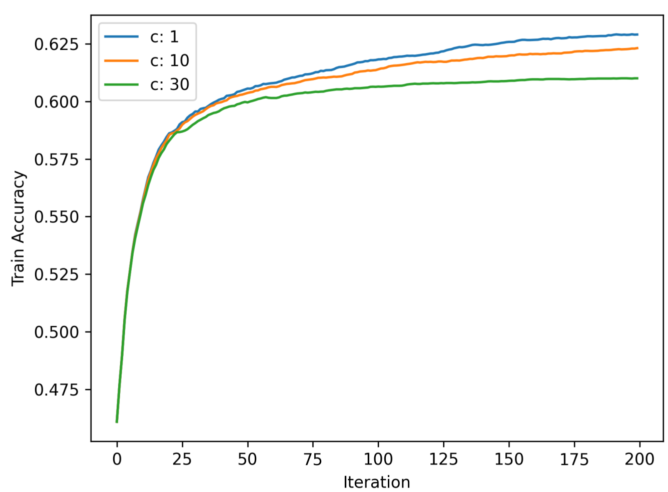

Figure 2: Training Accuracy Plot

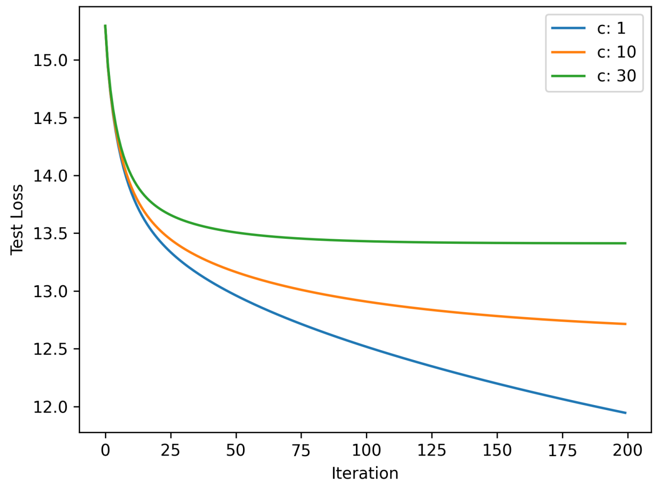

Figure 3: Test Loss Plot

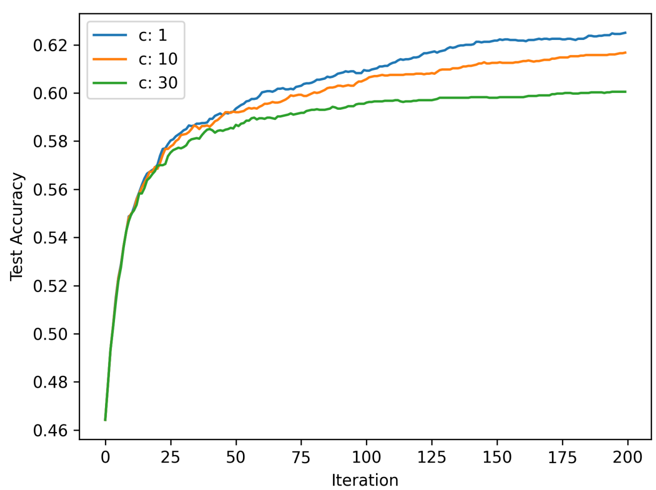

Figure 4: Test Accuracy Plot