IFT-6390 Fundamentals of Machine Learning
Professor: Ioannis Mitliagkas
Students: Abhay Puri (20209505), Saurabh Bodhe (20208545)

Homework 2 - Practical Report

1.

$$
\begin{aligned}
r(w) &= \frac{C}{2}\sum_{j'=1}^{m} ||w^{j'}||^2 \\
&= \frac{C}{2}\sum_{j'=1}^{m}\left(\sqrt{\sum_{k=1}^{p} {w_k^{j'}}^2}\right)^2 \\
&= \frac{C}{2}\sum_{j'=1}^{m}\sum_{k=1}^{p} {w_k^{j'}}^2 \\
\frac{\partial r}{\partial w_k^j} &= C w_k^j
\end{aligned}
$$

2.

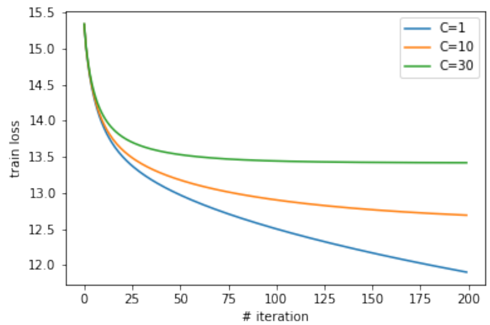
$$
\begin{aligned}
f(w,S) &= \frac{1}{n}\sum_{(x_i,y_i)\in S}\sum_{j'=1}^{m} \mathcal{L}(w^{j'};(x_i,y_i)) \\
\frac{\partial f}{\partial w_k^j} &= \frac{1}{n}\sum_{(x_i,y_i)\in S}\frac{\partial}{\partial w_k^j}\Big(max(0, 2-\langle w^j, x_i\rangle \mathbb{1}\{y_i=j\}\Big)^2
\end{aligned}
$$

$$
= \frac{2}{n}\sum_{(x_i,y_i)\in S}\Big(max(0, 2-\langle w^j, x_i\rangle \mathbb{1}\{y_i=j\}\Big)\mathcal{I}\{2-\langle w^j, x_i\rangle \mathbb{1}\{y_i=j\}>0\}(-x_{i,k}\mathbb{1}\{y_i=j\})
$$

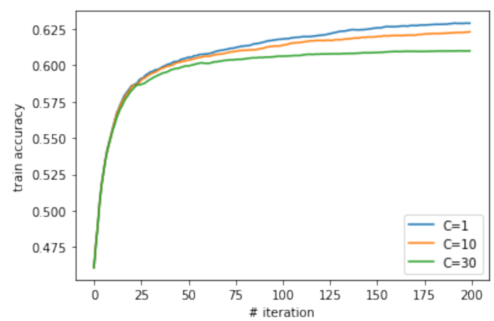
When $2-\langle w^j, x_i\rangle \mathbb{1}\{y_i=j\}<0$ the indicator function $\mathcal{I}$ returns 0. But even the max function returns 0 in this case. So the indicator function is redundant here.

$$
= \frac{-2}{n}\sum_{(x_i,y_i)\in S}\Big(max(0, 2-\langle w^j, x_i\rangle \mathbb{1}\{y_i=j\}\Big)\, x_{i,k}\mathbb{1}\{y_i=j\}
$$

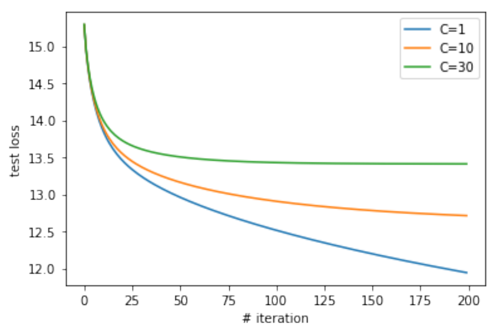
4. SVM plots



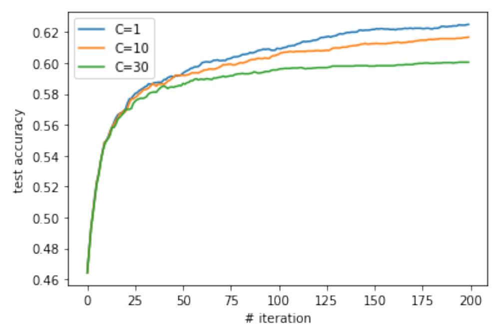
(a) Training loss



(b) Training accuracy



(c) Testing loss



(d) Testing accuracy

Figure 1: Loss and accuracy plots for different values of C

Derivation of Q2 (Detailed) :-

$$f(w,S) = \frac{1}{n} \sum_{(x_i,y_i)\in S} \sum_{j'=1}^{m} h\left(w^j ; (x_i, y_i)\right)$$

$$\frac{\partial f}{\partial w_k^j} = \frac{1}{n} \sum_{(x_i,y_i)\in S} \frac{\partial h}{\partial w_k^j} \left(w^j ; (x_i, y_i)\right)$$

$$= \frac{1}{n} \sum_{(x_i,y_i)\in S} \frac{\partial}{\partial w_k^j} \left(\max\left(0, 2 - \langle w^j, x_i \rangle 1\{y_i = j\}\right)\right)^2$$

Let $a = 2 - \langle w^j, x_i \rangle 1\{y_i = j\}$

& $m = \max(0, a)$

As we know, $\frac{\partial m}{\partial a} = \mathbb{I}(a > 0)$

$$\frac{\partial f}{\partial w_k^j} = \frac{1}{n} \sum_{(x_i,y_i)\in S} \frac{\partial}{\partial w_k^j} m^2 = \frac{1}{n} \sum_{(x_i,y_i)\in S} (2m) \frac{\partial m}{\partial w_k^j}$$

$$= \frac{2}{n} \sum_{(x_i,y_i)\in S} m \frac{\partial m}{\partial w_k^j} = \frac{2}{n} \sum_{(x_i,y_i)\in S} m \frac{\partial m}{\partial a} \frac{\partial a}{\partial w_k^j}$$

$$\frac{\partial a}{\partial w_k^j} = - x_{i,k} 1\{y_i = j\}$$

$$\Rightarrow \frac{\partial f}{\partial w_k^j} = \frac{2}{n} \sum_{(x_i,y_i)\in S} m\, \mathbb{I}(a > 0) \left(-x_{i,k} 1\{y_i = j\}\right)$$

when $a <= 0$, $\mathbb{I}(a>0) = 0$ but also $m = 0$ $\therefore \mathbb{I}(a>0)$ is redundant

Answer

$$\Rightarrow \frac{\partial f}{\partial w_k^j} = -\frac{2}{n} \sum_{(x_i,y_i)\in S} \left[\max\left(0, 2 - \langle w^j, x_i \rangle 1\{y_i = j\}\right) * x_{i,k} 1\{y_i = j\}\right]$$