Data Science
Yonsei University
Fall 2023
Kim, Songkuk

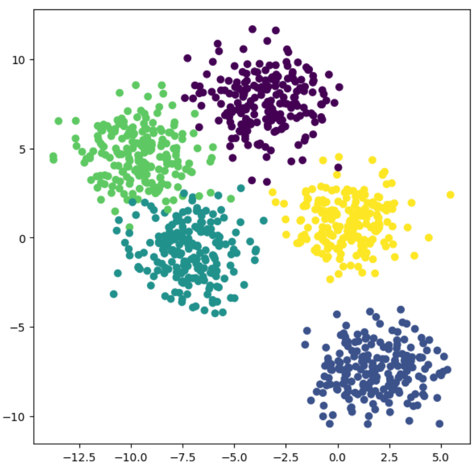
Homework 2

Implement multi-class logistic regression from scratch.

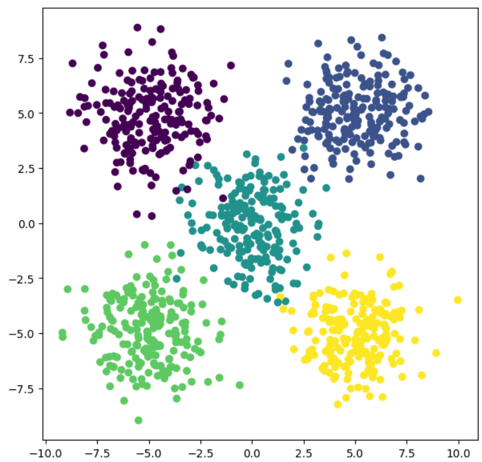
1. Do not use complex library functions except for generating test data.
2. Avoid loop and list comprehension.
3. Conduct experiments including the following data.

```
n_clusters = 5
n_samples = n_clusters * 200

#Generate dataset 1
X, y_true = make_blobs(
    n_samples=n_samples, centers=n_clusters, cluster_std=1.5,
    random_state=2023
)
#Generate dataset 2
X, y_true = make_blobs(
    n_samples=n_samples,
    centers=((-5, 5), (5, 5), (0, 0), (-5, -5), (5, -5)),
    cluster_std=1.5, random_state=2023
)
```



(a) Dataset 1



(b) Dataset 2

4. Use 60% data for training and 40% for test.

Gradient of loss

Let $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_N \end{bmatrix}$, $\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_N \end{bmatrix}$, $\mathcal{S} = \begin{bmatrix} \mathcal{S}_1 \\ \mathcal{S}_2 \\ \vdots \\ \mathcal{S}_N \end{bmatrix}$, $\boldsymbol{W} = \begin{bmatrix} | & | & & | \\ \boldsymbol{w_1} & \boldsymbol{w_2} & \cdots & \boldsymbol{w_C} \\ | & | & & | \end{bmatrix}$

$$p(y = c|\boldsymbol{x}_n, \boldsymbol{W}) = \mathcal{S}_{n,c} = \frac{e^{\boldsymbol{a}_{n,c}}}{\sum_{k=0}^{C} e^{\boldsymbol{a}_{n,k}}}$$

where $a_{n,k} = \boldsymbol{x}_n \boldsymbol{w}_k$ and $\mathcal{S}_{n,c} = softmax(a_{n,c}; a_{n,1}, \cdots, a_{n,C})$.

$$\begin{aligned} NLL(\boldsymbol{W}) &= -\frac{1}{N} \log \prod_{n=1}^{N} \prod_{c=1}^{C} \mathcal{S}_{n,c}^{y_{n,c}} \\ &= -\frac{1}{N} \sum_{n=1}^{N} \sum_{c=1}^{C} y_{n,c} \log \mathcal{S}_{n,c} \\ &= -\frac{1}{N} \mathbb{H}(\boldsymbol{Y}, \boldsymbol{\mathcal{S}}) \end{aligned}$$

where $\mathcal{S}_{n,c} = \mathcal{S}_c(\boldsymbol{x}_n)$.

Gradient $g(\boldsymbol{W}) = \frac{1}{N} \sum_n \nabla_{\boldsymbol{W}} NLL_n$
For brevity, let $s_c = \mathcal{S}_{n,c},\ a_j = \boldsymbol{x}_n \boldsymbol{w}_j,\ y_c = y_{n,c}$

$$\begin{aligned} \nabla_{\boldsymbol{w}_j} NLL_n &= \sum_{c=1}^{C} \frac{\partial NLL_n}{\partial s_c} \frac{\partial s_c}{\partial a_j} \frac{\partial a_j}{\partial \boldsymbol{w}_j} \\ &= -\sum_{c=1}^{C} y_c (I_{cj} - s_j) \boldsymbol{x}_n \\ &= (s_j - y_j) \boldsymbol{x}_n \end{aligned}$$

$$\because \frac{\partial NLL_n}{\partial s_c} = -\frac{y_c}{s_c},\ \frac{\partial s_c}{\partial a_j} = s_c(I_{cj} - s_j),\ \frac{\partial a_j}{\partial \boldsymbol{w}_j} = \boldsymbol{x}_n$$

$$g_n(\boldsymbol{W}) = \begin{bmatrix} | & | & & | \\ (s_1 - y_1)\boldsymbol{x}_n & (s_2 - y_2)\boldsymbol{x}_n & \cdots & (s_C - y_C)\boldsymbol{x}_n \\ | & | & & | \end{bmatrix} = \boldsymbol{x}_n^\top (\mathcal{S}_n - \boldsymbol{y}_n)$$

$$g(\boldsymbol{W}) = \frac{1}{N} \sum_{n=1}^{N} g_n(\boldsymbol{W}) = \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{x}_n^\top (\mathcal{S}_n - \boldsymbol{y}_n)$$