學號：R07521603 系級：土木所碩二 姓名：蔡松霖

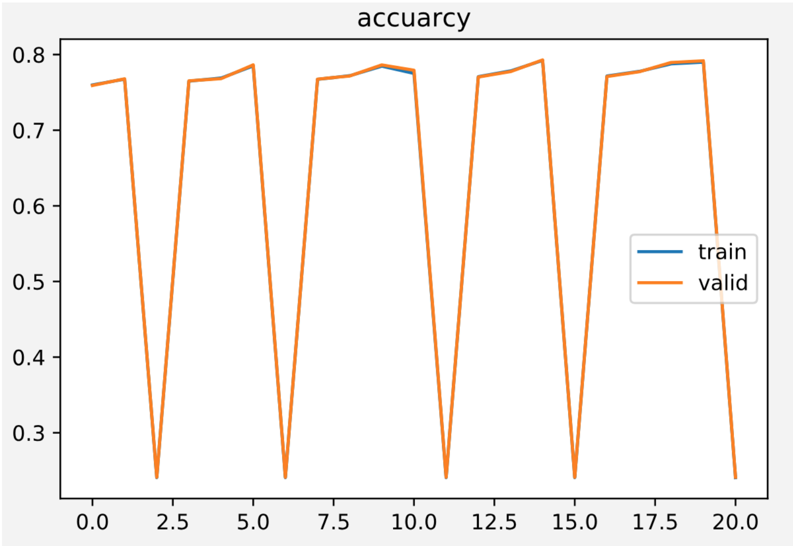
1. (0.5%) 請比較你實作的 generative model、logistic regression 的準確率，何者較佳?

	public leaderboard score	private leaderboard score
generative model	0.84434	0.84412
logistic regression	0.85503	0.85014

不論是 public score 還是 private score，皆是 logistic regression 的準確率較佳。

2. (0.5%) 請實作特徵標準化(feature normalization)並討論其對於你的模型準確率的影響

這邊同樣以 logistic regression model 作為 model 測試標準化(減去 mean，除以 std)的影響。可以觀察到有沒有實作特徵標準化對於 training 有很顯著的影響，沒有標準化基本上沒辦法 train 下去，training 過程一直大幅度的上下震盪，且準確度沒有突破 0.8 如下圖。



Model 準確率的部分，沒有標準化時 training score 無法突破 0.8，最後在 kaggle leaderboard 的 score 整理如下表所示：

	public leaderboard score	private leaderboard score
有實作特徵標準化	0.85503	0.85014
沒實作特徵標準化	0.23525	0.23719

3. (1%) 請說明你實作的 best model，其訓練方式和準確率為何?

我使用 scikit-learn 的 GradientBoostingClassifier 來實作 best model，data preprocessing 的部分使用 min_max_scaler 來特徵標準化。基本上就是進行參數 grid search 調參找 cross validation score 較佳的 model。最後選擇的參數如下：

GradientBoostingClassifier(criterion='friedman_mse', init=None,
learning_rate=0.025, loss='deviance', max_depth=18, max_features='sqrt',

max_leaf_nodes=None, min_impurity_decrease=0.0, min_impurity_split=None, min_samples_leaf=20, min_samples_split=200, min_weight_fraction_leaf=0.0, n_estimators=500, n_iter_no_change=None, presort='auto', random_state=42, subsample=0.8, tol=0.0001, validation_fraction=0.1, verbose=0, warm_start=True)

Public score: 0.86363，private score: 0.85714

4. (3%) Refer to math problem
https://hackmd.io/0fDimqO7RaSCPpD_minSGQ?both

1.
ref: https://people.eecs.berkeley.edu/~jrs/189/exam/mids14.pdf
ref: https://stats.stackexchange.com/questions/125557/deriving-the-maximum-likelihood-for-a-generative-classification-model-for-k-clas
ref: https://math.stackexchange.com/questions/421105/maximum-likelihood-estimator-of-parameters-of-multinomial-distribution

1. $P(C_k) = \pi_k$.

$$p(x,t) = p(x|t)\,p(t) = \prod_{k=1}^{K} \left(p(x|C_k)\,\pi_k \right)^{t_k}$$

denote the parameters as θ.

$$L(\theta) = \prod_{n=1}^{N} \prod_{k=1}^{K} \left(p(x|C_k)\,\pi_k \right)^{t_{n,k}}$$

取 ln

$$\Rightarrow l(\theta) = \sum_{n=1}^{N} \sum_{k=1}^{K} t_{n,k} \left[\ln p(x|C_k) + \ln \pi_k \right]$$

posing a constraint ($\sum_{k=1}^{K} \pi_k = 1$) with Lagrange multiplier

$$\mathcal{L}(\pi, \lambda) = \sum_{n=1}^{N} \sum_{k=1}^{K} t_{n,k} \left[\ln p(x|C_k) + \ln \pi_k \right] + \lambda \left(1 - \sum_{k=1}^{K} \pi_k \right)$$

$$\frac{\partial}{\partial \pi_k} \mathcal{L}(\pi, \lambda) = \frac{1}{\pi_k} \sum_{n=1}^{N} t_{n,k} - \lambda = 0 \Rightarrow \pi_k = \frac{1}{\lambda} \sum_{n=1}^{N} t_{n,k} = \frac{N_k}{\lambda}$$

$$\frac{\partial}{\partial \lambda} \mathcal{L}(\pi, \lambda) = 1 - \sum_{k=1}^{K} \pi_k = 0 \Rightarrow \sum_{k=1}^{K} \pi_k = 1$$

$$\sum_{k=1}^{K} \pi_k = \sum_{k=1}^{K} \frac{N_k}{\lambda} = \frac{N}{\lambda} = 1 \Rightarrow \lambda = N$$

$$\therefore \pi_k = \frac{N_k}{N} \ *$$

2.
ref: https://people.eecs.berkeley.edu/~jordan/courses/260-spring10/other-readings/chapter13.pdf
ref: https://en.wikipedia.org/wiki/Adjugate_matrix

2. $$\frac{\partial \log(\det\Sigma)}{\partial \sigma_{ij}} = \frac{1}{|\Sigma|}\frac{\partial}{\partial\sigma_{ij}}|\Sigma|$$

$$|\Sigma| = \left((-1)^{i+j}\sigma_{ij} M_{ij}\right)_{1\le i,j\le m}$$

$$\frac{\partial}{\partial\sigma_{ij}}|\Sigma| = \left((-1)^{i+j}M_{ij}\right)_{1\le i,j\le m} = C = \mathrm{adj}(\Sigma)^T$$

By $\boxed{A^{-1} = \frac{1}{\det(A)}\mathrm{adj}(A)}$

$$\Rightarrow \Sigma^{-1} = \frac{1}{|\Sigma|}\mathrm{adj}(\Sigma) = \frac{1}{|\Sigma|}\left((-1)^{i+j}M_{ji}\right)_{1\le i,j\le m}$$

$$\Rightarrow e_j\Sigma^{-1}e_i^T = \frac{1}{|\Sigma|}\left((-1)^{i+j}M_{ij}\right)_{1\le i,j\le m} = \frac{1}{|\Sigma|}\frac{\partial}{\partial\sigma_{ij}}|\Sigma| = \frac{\partial\log|\Sigma|}{\partial\sigma_{ij}}$$ #

M_{ij}: the determinant of the $(m-1)\times(m-1)$ matrix that results from deleting row i and column j of A.

C: the cofactor matrix of A

$$C = \left((-1)^{i+j}M_{ij}\right)_{1\le i,j\le m}$$

$$\mathrm{adj}(A) = C^T = \left((-1)^{i+j}M_{ji}\right)_{1\le i,j\le m}$$

3.
ref: https://people.eecs.berkeley.edu/~jordan/courses/260-spring10/other-readings/chapter13.pdf
ref: https://github.com/zhengqigao/PRML-Solution-Manual

logarithm likelihood.

3. 同1: $$\ln p = \sum_{n=1}^{N}\sum_{k=1}^{K} t_{n,k}\left[\ln p(x_n|C_k) + \ln\pi_k\right]$$

provided $p(x|C_k) = \mathcal{N}(x|\mu_k,\Sigma)$,

$$\ln p \propto \sum_{n=1}^{N}\sum_{k=1}^{K} t_{n,k}\left[-\frac12\ln|\Sigma| - \frac12(x_n-\mu_k)\Sigma^{-1}(x_n-\mu_k)^T\right]$$

$$\frac{\partial \ln p}{\partial\mu_k} = \sum_{n=1}^{N} t_{n,k}\Sigma^{-1}(x_n-\mu_k) \overset{set}{=} 0$$

$$\Rightarrow \sum_{n=1}^{N} t_{n,k}\Sigma^{-1}x_n = \sum_{n=1}^{N} t_{n,k}\Sigma^{-1}\mu_k = N_k\Sigma^{-1}\mu_k$$

同乘 Σ/N_k

$$\Rightarrow \mu_k = \frac{1}{N_k}\sum_{n=1}^{N} t_{n,k}x_n$$ #

$$\frac{\partial\ln p}{\partial\Sigma} = \sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk}\left(-\frac12\Sigma^{-1}\right) - \frac12\frac{\partial}{\partial\Sigma}\sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk}(x_n-\mu_k)\Sigma^{-1}(x_n-\mu_k)^T$$

$$= \sum_{n=1}^{N}\sum_{k=1}^{K} -\frac{t_{nk}}{2}\Sigma^{-1} - \frac12\frac{\partial}{\partial\Sigma}\sum_{k=1}^{K}\sum_{n=1}^{N} t_{nk}(x_n-\mu_k)\Sigma^{-1}(x_n-\mu_k)^T$$

$$= \sum_{n=1}^{N} -\frac12\Sigma^{-1} - \frac12\frac{\partial}{\partial\Sigma}\sum_{k=1}^{K} N_k\,\mathrm{Trace}(\Sigma^{-1}S_k) \quad \left(S_k = \frac{1}{N_k}\sum_{n=1}^{N} t_{nk}(x_n-\mu_k)(x_n-\mu_k)^T\right)$$

$$= -\frac N2\Sigma^{-1} + \frac12\sum_{k=1}^{K} N_k\Sigma^{-1}S_k\Sigma^{-1} \overset{set}{=} 0$$

$$\Rightarrow \Sigma = \sum_{k=1}^{K}\frac{N_k}{N}S_k$$ #