

## A PROOFS AND DERIVATIONS

PROOF 1. (**Proposition 1**) For any two sub-matrix :  $X_1, X_2 \in R^{N \times 1}$ , regardless of row or column sub-matrix of the original matrix. Note for any matrix  $X_S$  with column = 1,  $(X_S^T X_S)^{-1} = \|X_S\|^2 = \text{Vol}(X_S)^2$ . Considering the pseudo-inverse is denoted as:  $X_S^+ := (X_S^T X_S)^{-1} X_S^T$ , the bias variation is:

$$\begin{aligned}
 \text{bias}_2^2 - \text{bias}_1^2 &= \|X^+ - X_2^+\|^2 - \|X^+ - X_1^+\|^2 \\
 &= \|X^+\|^2 - 2 \langle X^+, X_2^+ \rangle + \|X_2^+\|^2 - \|X^+\|^2 + 2 \langle X^+, X_1^+ \rangle - \|X_1^+\|^2 \\
 &= \|X^+\|^2 - 2 \|X_2^+\|^2 + \|X_2^+\|^2 - \|X^+\|^2 + 2 \|X_1^+\|^2 - \|X_1^+\|^2 \\
 &= \|X_1^+\|^2 - \|X_2^+\|^2 = \frac{1}{\|X_1\|^4} \|X_1^T\|^2 - \frac{1}{\|X_2\|^4} \|X_2^T\|^2 \\
 &== \frac{1}{\|X_1\|^2} - \frac{1}{\|X_2\|^2} = \frac{1}{\text{Vol}(X_1)^2} - \frac{1}{\text{Vol}(X_2)^2}
 \end{aligned} \tag{1}$$

Since the  $\text{Vol}(X)$  is constant under  $\text{column} = 1$ , a larger volume yields a smaller bias is proved. Combined with the relation  $\text{variation}_1 - \text{variation}_2 = \text{bias}_2 - \text{bias}_1$ , a larger volume yields a larger parameter variation is apparent.

PROOF 2. (**Proposition 2**) The proposition 2 is expanded by substituting  $X^+ = G^{-1} X^T$ , where  $G := X^T X$  is the left Gram matrix.

$$\begin{aligned}
 \text{bias}_2^2 - \text{bias}_1^2 &:= \|X^+ - X_2^+\|^2 - \|X^+ - X_1^+\|^2 \\
 &= \|X^+\|^2 - 2 \langle X^+, X_2^+ \rangle + \|X_2^+\|^2 - \|X^+\|^2 + 2 \langle X^+, X_1^+ \rangle - \|X_1^+\|^2 \\
 &= \|X_2^+\|^2 - \|X_1^+\|^2 + 2 \langle X^+, X_1^+ - X_2^+ \rangle \\
 &= \|G_2^{-1} X_2^T\|^2 - \|G_1^{-1} X_1^T\|^2 + 2 \langle G^{-1} X^T, G_1^{-1} X_1^T - G_2^{-1} X_2^T \rangle
 \end{aligned} \tag{2}$$

Inspired by the work of Xu et al.[1],  $G^{-1}$  can be expanded by Sylvester's matrix theorem.

THEORY 1. (**Sylvester's Matrix Theorem**). Given a diagonalizable square matrix  $X$ , any analytic function  $f()$  can be expanded,

$$f(X) = \sum_{l=1}^k f(\lambda_l) X_l, \tag{3}$$

where  $\lambda_l$  is the  $l$ -th distinct eigenvalue of  $X$  and  $X_l$  is the Frobenius covariant as follows:

$$X_l := \prod_{j=1, j \neq l}^k \frac{1}{\lambda_l - \lambda_j} (X - \lambda_j I). \tag{4}$$

COROLLARY 1. Supposing  $f()$  is a inverse function, then,

$$f(X) = X^{-1} = \sum_{l=1}^k \frac{1}{\lambda_l} X_l \tag{5}$$

For  $G$  is diagonalizable matrix,  $G^{-1}$  can be written as follows by above corollary.

$$G^{-1} = \sum_{l=1}^k \frac{1}{\lambda_l} G_l = \prod_{j=1, j \neq l}^k \frac{1}{\lambda_l - \lambda_j} \prod_{j=1, j \neq l}^k (G - \lambda_j I) \tag{6}$$

Denoting  $p_l = \prod_{j=1, j \neq l}^k \frac{1}{\lambda_l - \lambda_j}$  and expanding the products, we have:

$$\frac{1}{p_l} = \underbrace{\Lambda \sum_{g=1}^k (-1)^{g+1} \lambda_l^{k-g} \left[ \sum_{\mathcal{H} \subseteq \{1, \dots, k\} \setminus \{l\}, |\mathcal{H}|=g-1} \left( \prod_{h \in \{1, \dots, k\} \setminus \mathcal{H}} \frac{1}{\lambda_h} \right) \right]}_{\sigma_l} \tag{7}$$

where  $\Lambda = \prod_{i=1}^k \lambda_i = |\mathbf{G}|$ . The Equation 6 can be rewritten as:

$$\begin{aligned} \mathbf{G}^{-1} &= \frac{1}{|\mathbf{G}|} \frac{1}{\sigma_l} \underbrace{\prod_{j=1, j \neq l}^k (\mathbf{G} - \lambda_j \mathbf{I})}_{\mathbf{Q}} \\ &= \frac{1}{\text{Vol}(\mathbf{X})^2} \frac{1}{\sigma_l} \underbrace{\prod_{j=1, j \neq l}^k (\mathbf{G} - \lambda_j \mathbf{I})}_{\mathbf{Q}} \end{aligned} \quad (8)$$

Substituting  $\mathbf{G}^{-1}(\mathbf{G}_1^{-1}, \mathbf{G}_2^{-1})$  in Equation 2 by above result, the Equation 2 can be rewritten as:

$$\text{bias}_2^2 - \text{bias}_1^2 = \frac{\|Q_2 X_2^T\|^2}{\text{Vol}(X_2)^4} - \frac{\|Q_1 X_1^T\|^2}{\text{Vol}(X_1)^4} + 2 \left\langle \frac{Q_X X^T}{\text{Vol}(X)^2}, \frac{Q_1 X_1^T}{\text{Vol}(X_1)^2} - \frac{Q_2 X_2^T}{\text{Vol}(X_2)^2} \right\rangle \quad (9)$$

LEMMA 1. For any matrix  $X = [X_1^T, X_2^T]^T \in \mathbb{R}^{N \times M}$  and  $X_1^T, X_2^T \in \mathbb{R}^{\frac{N}{2} \times M}$  are submatrices in row, there exists  $\text{Vol}(X) \gg \max(\text{Vol}(X_1), \text{Vol}(X_2))$  when  $N \rightarrow \infty$ . The relation is same for column submatrices.

LEMMA 2. For a invertible matrix  $A$  and any column vector  $u$  and  $v$ , then there exists:

$$\det(A + uv^T) = \det(A) (1 + v^T A^{-1} u) \quad (10)$$

PROOF 3. (**Lemma 1**) Considering a matrix  $\tilde{X} = [X^T, x^T]^T$ , where  $X \in \mathbb{R}^{n \times m}$  is a submatrix,  $x \in \mathbb{R}^{1 \times m}$  is a row vector. The square volume of  $\tilde{X}$  can be written as:

$$\begin{aligned} \text{Vol}(\tilde{X})^2 &= |\tilde{X}^T \tilde{X}| = \left| \begin{bmatrix} X^T & x^T \end{bmatrix} \begin{bmatrix} X \\ x \end{bmatrix} \right| \\ &= |X^T X + x^T x| = \sigma_1 |X^T X|. \end{aligned} \quad (11)$$

According to the property of matrix determinant lemma (Lemma 2) and  $x(X^T X)^{-1} x^T > 0$ , the constant coefficient  $\sigma = 1 + x(X^T X)^{-1} x^T$  is greater than 1. Considering another submatrix  $X' = [x_1^T, \dots, x_n^T]^T$  and adding its row vectors in matrix  $X$  row by row, the square volume changing is written as:

$$\begin{aligned} \text{Vol}([X^T, X'^T]^T) &= \text{Vol}([X^T, x_1^T, \dots, x_n^T]^T) \\ &= \sigma_n \text{Vol}([X^T, x_1^T, \dots, x_{n-1}^T]^T) = \dots = \prod_{i=1}^n \sigma_i \text{Vol}(X), \end{aligned} \quad (12)$$

where  $\sigma_i = 1 + x_i(X^T X)^{-1} x_i^T$ . For each  $\sigma_i$  is greater than 1,  $\prod_{i=1}^n \sigma_i \rightarrow \infty$  when  $n \rightarrow \infty$ . Above all, **Lemma 1** is proved taking  $X_1$  and  $X_2$  as a new addition submatrix, respectively. Same conclusions can be applied in column submatrices.

PROOF 4. (**Proposition 3**) Considering a replication-involving dataset  $X_{rep} = \begin{bmatrix} X^T & \underbrace{X_S^T, \dots, X_S^T}_d \end{bmatrix}^T$ , where  $X_S$  is row vector and is replicated for  $d$  times. According to Equation 12, the square volume of  $X_{rep}$  is written as:

$$\text{Vol}(X_{rep})^2 = (1 + X_S(X^T X)^{-1} X_S^T)^d |X^T X| \quad (13)$$

For  $(1 + X_S(X^T X)^{-1} X_S^T) > 1$ , the exponential increasing of volume under replication is proved and  $\lim_{d \rightarrow \infty} \text{Vol}(X_{rep}) = \infty$  is hold. The same result can be applied in row submatrices.

PROOF 5. (**Proposition 5**) Considering a replication-involving matrix  $X_{rep} = \text{replicate}(X, c)$ , the inflation is written as:

$$\text{inflation} = \frac{RV_{\text{cluster}}(X_{rep})}{RV_{\text{cluster}}(X)} = \frac{\text{Vol}(\tilde{X}_{rep}) \prod_{i \in K} \rho_{rep,i}}{\text{Vol}(\tilde{X}) \prod_{i \in K} \rho_i} \quad (14)$$

Due to direct copying, the clusters in  $X_{rep}$  and  $X$  are with similar centers, thus,  $\text{Vol}(\tilde{X}_{rep}) \approx \text{Vol}(\tilde{X})$ .

Following summation formula of geometric progression, when  $0 < \alpha < 1$ ,

$$1 \leq \rho_{rep,i} := \sum_{p=0}^{\phi_{rep,i}} \alpha^p = \frac{1 - \alpha^{(\phi_{rep,i}+1)}}{1 - \alpha} \leq \frac{1}{1 - \alpha} \quad (15)$$

$$1 \leq \rho_i := \sum_{p=0}^{\phi_i} \alpha^p = \frac{1 - \alpha^{(\phi_i+1)}}{1 - \alpha} \leq \frac{1}{1 - \alpha} \quad (16)$$

Thus,

$$(1 - \alpha)^K \leq \text{inflation} \leq (1 - \alpha)^{-K} \quad (17)$$

PROOF 6. (**Proposition 6**) When  $\alpha = 1/\beta N$ , the following inequality relation exists:

$$1 \leq \rho_{rep,i} = \sum_{p=0}^{\phi_{rep,i}} \frac{1}{\beta N} \alpha^p \leq \frac{1}{1 - \frac{1}{\beta N}} = 1 + \frac{1}{\beta N - 1} \quad (18)$$

$$1 \leq \rho_i = \sum_{p=0}^{\phi_i} \frac{1}{\beta N} \alpha^p \leq \frac{1}{1 - \frac{1}{\beta N}} = 1 + \frac{1}{\beta N - 1} \quad (19)$$

Combined with Equation 14,

$$\left(1 + \frac{1}{\beta N - 1}\right)^{-K} \leq \text{inflation} \leq \left(1 + \frac{1}{\beta N - 1}\right)^K \quad (20)$$

When  $N \rightarrow \infty$  and  $K$  is pre-determined, the following limit theorem is exist.

$$\lim_{N \rightarrow \infty} \left(1 + \frac{1}{\beta N - 1}\right)^{-K} = \lim_{N \rightarrow \infty} \left(1 + \frac{1}{\beta N - 1}\right)^K = 1 \quad (21)$$

Thus,  $\text{inflation} \rightarrow 1$  is hold under  $N \rightarrow \infty$ .

PROOF 7. (**Proposition 7**) Taking the scenario where the dataset  $X$  is clustered to  $K$  partitions as an example. The matrix of cluster centers is denoted as  $\tilde{X}$ . For a balanced dataset, each clusters  $C_i, i = 1, \dots, K$  is supposed to contain  $D$  data points. Then, we have

$$\begin{aligned} \text{Vol}(X_1)^2 &= \left| \left[ (\tilde{X}_1 + \Gamma_1)^T, \dots, (\tilde{X}_1 + \Gamma_D)^T \right]^T \begin{bmatrix} \tilde{X}_1 + \Gamma_1 \\ \vdots \\ \tilde{X}_1 + \Gamma_1 \end{bmatrix} \right| \\ &= \left| \sum_{i=1}^D (\tilde{X}_1 + \Gamma_1)^T (X_1 + \Gamma_i) \right| \end{aligned} \quad (22)$$

where  $\tilde{X}_1$  is the matrix of cluster centers and  $\Gamma_i = \begin{bmatrix} C_{0,i} - \tilde{X}_0 \\ \vdots \\ C_{K,i} - \tilde{X}_K \end{bmatrix}$  denotes the relative distance of  $i$ -th data points in each clusters to its cluster's center. Each cluster center generated by  $k$ -Means is:

$$\tilde{X}_{1,k} = \frac{\sum_{i=1}^D \delta_{ik} x_i}{\sum_{i=1}^D \delta_{ik}}. \quad (23)$$

where  $\delta_{ik}$  is a cluster indicator variable with  $\delta_{ik} = 1$  if  $x_i$  in  $k$ -th cluster. As the number of data points increases, the distance vectors of the points to cluster center are cancel each other out,

$$\sum_{i=0}^D \Gamma_D \rightarrow \vec{0}. \quad (24)$$

Thus, the Equation 22 can be rewritten as:

$$\text{Vol}(X_1)^2 = \left| \sum_{i=1}^D (\tilde{X}_1)^T (\tilde{X}_1) + \sum_{i=1}^D (\tilde{\Gamma}_i)^T (\Gamma_i) \right|, \quad (25)$$

In the same way to get

$$\text{Vol}(X_2)^2 = \left| \sum_{i=1}^D (\tilde{X}_2)^T (\tilde{X}_2) + \sum_{i=1}^D (\tilde{\Upsilon}_i)^T (\Upsilon_i) \right| \quad (26)$$

Applying determinant property,

$$\begin{aligned} \frac{\text{Vol}(X_1)^2}{\text{Vol}(X_2)^2} &= \frac{\left| D \cdot (\tilde{X}_1)^T (\tilde{X}_1) + \sum_{i=1}^D (\Gamma_i)^T (\Gamma_i) \right|}{\left| D \cdot (\tilde{X}_2)^T (\tilde{X}_2) + \sum_{i=1}^D (\Upsilon_i)^T (\Upsilon_i) \right|} \\ &= \frac{\left| (\tilde{X}_1)^T (\tilde{X}_1) + \frac{1}{D} \sum_{i=1}^D (\Gamma_i)^T (\Gamma_i) \right|}{\left| (\tilde{X}_2)^T (\tilde{X}_2) + \frac{1}{D} \sum_{i=1}^D (\Upsilon_i)^T (\Upsilon_i) \right|} \end{aligned} \quad (27)$$

When  $N \rightarrow \infty$ , the data points in each cluster  $D \rightarrow \infty$ , the Equation 27 can be rewritten as:

$$\frac{\text{Vol}(X_1)^2}{\text{Vol}(X_2)^2} = \frac{|\tilde{X}_1^T \tilde{X}_1|}{|\tilde{X}_2^T \tilde{X}_2|} = \frac{\text{Vol}(\tilde{X}_1)^2}{\text{Vol}(\tilde{X}_2)^2} \quad (28)$$

When  $N \rightarrow \infty$ , proposition 6 has proved the following relation,

$$\frac{RV(\tilde{X}_1)}{RV(\tilde{X}_2)} \rightarrow \frac{\text{Vol}(\tilde{X}_1)}{\text{Vol}(\tilde{X}_2)} \quad (29)$$

Combined Equation 29 with Equation 28,  $RV(\tilde{X}_1)/RV(\tilde{X}_2)/[V(X_1)/V(X_2)] \rightarrow 1$  under  $N \rightarrow \infty$  is hold.

**DEFINITION 1. (Unbounded subset-sum problem)** Given a set of positive integers  $\{k_0, \dots, k_n\}$ , an unbounded subset-sum problem is defined as to find the non-negative integers  $\alpha_i$  so that  $\sum_{i=1}^n \alpha_i k_i = K$ , for we can achieve  $K$  by  $k_i$  for any times, it's known that unbounded subset-sum problem is NP-hard.

**LEMMA 3.** Let  $v_i = p_i, i = 1, \dots, n, p_{n+1} = K + \Delta$  when  $v_{n+1} = K$  and  $\Delta \in (0, 1)$ , a subadditive and monotone function  $p(x)$  interpolating on the points  $(v_i, p_i)$  exist if and only if unbounded subset sum  $\sum_{i=1}^n \alpha_i v_i = K$  not exists.

**PROOF 8. (Lemma 3)** If  $\sum_{i=1}^n \alpha_i p_i = K$  exists, then we have:

$$K + \Delta = p(K) = p\left(\sum_{i=1}^n \alpha_i v_i\right) \leq \sum_{i=1}^n \alpha_i p_i \stackrel{v_i = p_i}{=} K \quad (30)$$

$K + \Delta = K$  is a contradiction so that if  $\sum_{i=1}^n \alpha_i v_i = K$  exists, a subadditive and monotone function  $p(x)$  interpolating on the  $n + 1$  points  $(v_i, p_i)$  is not exist.

Conversely, in the next, we prove if  $\sum_{i=1}^n \alpha_i p_i = K$  not exists, we can construct a subadditive and monotone function  $p(x)$  that interpolates the  $(n + 1)$  points. First, we introduce a function  $(x)$  to reflect the smallest possible unbounded subset sum at  $x$ .  $(x)$  at least constrains an unbounded subset sum contains  $x$ , thus  $(x) \geq x$ . Then, we define a function  $p(x) = \min((x), K + \Delta)$  and our goal is to prove such  $p(x)$  is satisfied subadditive, monotone and interpolating on the  $n + 1$  points  $(v_i, p_i)$ . It is apparent that  $p(x)$  is monotone. Since a set containing  $x$ -self is a minimum unbounded subset sum, we have  $i = v_i = p_i \leq K + \Delta$ . For we have assumed that  $\sum_{i=1}^n \alpha_i p_i = K$  is not exist, thus  $(v_{i+1}) \geq K + 1$ . Then the  $p(x)$  can be written as:

$$p(x) = \begin{cases} \mu(x), & \mu(x) \leq K \\ K + \Delta, & \mu(x) \geq K + 1 \end{cases} \quad (31)$$

If  $\mu(x) \geq K + 1$ , then  $p(x + y) \leq K + \Delta = p(x) \leq p(x) + p(y)$ . The relation is also holds when  $\mu(y) \geq K + 1$ . When both  $\mu(x) \leq K$  and  $\mu(y) \leq K$ , we have  $p(x) = (x) = \sum_{i=1}^n \alpha_i v_i$  and  $p(y) = (y) = \sum_{i=1}^n \beta_i v_i$ . Then,  $x + y \leq p(x) + p(y) = \sum_{i=1}^n (\alpha_i + \beta_i) v_i$ . According to the definition of  $(x)$ ,  $p(x + y) = (x + y) = \min(x + y, \sum_{i=1}^n \gamma_i v_i) \leq \sum_{i=1}^n (\alpha_i + \beta_i) v_i = p(x) + p(y)$ .

Above all, we have proved Lemma 3). However, the unbounded subset-sum problem is NP-hard, whether the sufficient and necessary conditions that unbounded subset sum  $\sum_{i=1}^n \alpha_i v_i = K$  not existing in Lemma 3) is a co-NP hard problem.

**PROOF 9. (Proposition 8)** For the pricing function  $p$  satisfies  $p(x)/x \geq p(y)/y$  when  $x \leq y$ , then, it must have:

$$\begin{aligned} \frac{p(x+y)}{x+y} &\leq \min\left(\frac{p(x)}{x}, \frac{p(y)}{y}\right) \Rightarrow \\ p(x+y) &\leq \min\left(p(x) + \underbrace{\frac{yp(x)}{x}}_{\geq p(y)}, p(y) + \underbrace{\frac{xp(y)}{y}}_{\geq p(x)}\right) \\ &\leq p(x) + p(y) \end{aligned} \quad (32)$$

Constraint  $p(x)/x \geq p(y)/y, x \leq y$  representing a subspace of sub-additivity constraint is proved.

## B ADDITIONAL EXPERIMENT RESULTS

**EXPERIMENT 1. (Higher Cluster Diversity vs. Higher Data Diversity)** In the MNIST handwritten digit classification problem, for example, subset 1 contains 100 randomly sampled handwritten pictures ranging from clusters labeled "0" to "9", while subset 2 contains 200 handwritten pictures sampled from clusters labeled "0" to "5". Compared to subset 1, subset 2 is with higher data diversity for containing more unique data and lower cluster diversify for sampling from 5 clusters. Figure 1 shows that a model trained on subset 1 is with better classification performance than subset 2, meaning that a higher cluster diversity is more valuable to indicate better learning performance.

Subset1 from the 64-dimensional digits dataset

```

0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9
0 3 5 5 6 5 0 5 8 9
8 4 1 7 7 3 5 1 0 0
2 2 7 8 2 0 1 2 6 3
3 7 3 3 4 6 6 6 4 9
1 5 0 3 5 2 8 2 0 0
1 7 6 3 2 1 7 4 6 3
1 3 3 1 7 6 8 4 3 1

```

Subset2 from the 64-dimensional digits dataset

```

0 1 2 3 4 0 1 1 3 4 0 1 2 3 4 0 0 4 1 3
1 0 0 2 2 2 0 1 2 3 3 3 3 4 4 1 0 1 2 0
0 1 3 2 1 4 3 1 3 1 4 3 1 4 0 3 1 4 2
2 2 4 4 0 0 1 2 3 4 0 1 2 3 4 0 1 2 3 4
0 0 4 1 3 1 0 0 2 2 1 0 1 2 3 3 3 4 4
1 0 2 1 0 0 1 3 2 1 4 3 1 4 3 1 4 0 3
4 4 4 1 2 1 4 4 0 0 1 2 3 4 0 1 2 3 4 0
4 2 3 4 0 0 4 1 3 1 0 0 2 2 2 0 4 2 3 3
3 3 4 4 1 0 2 2 0 0 1 3 2 1 4 3 1 4 2
3 1 4 0 3 1 4 4 2 2 2 4 4 0 3 0 1 2 3 4
0 1 1 3 4 0 1 2 3 4 0 0 4 1 3 1 0 0 1 2
2 0 1 2 3 3 3 3 4 4 1 0 1 2 0 0 1 3 2 1
4 3 1 3 1 4 3 1 4 0 3 1 4 4 2 2 4 4 0
0 1 2 3 4 0 1 2 3 4 0 1 2 3 4 0 0 4 1 3
4 0 0 1 2 2 0 1 1 3 3 3 3 4 4 1 0 1 2 0
0 1 3 2 1 4 3 1 3 1 4 3 1 4 0 3 1 4 2
2 1 4 4 0 0 1 2 3 4 0 1 2 3 4 0 1 2 3 4
0 0 4 1 3 1 0 0 2 2 2 0 1 2 3 3 3 4 4
1 0 2 1 0 0 1 3 2 1 4 3 1 4 3 1 4 0
3 1 4 4 2 2 2 4 4 0 0 1 2 3 4 0 1 2 3 4

```

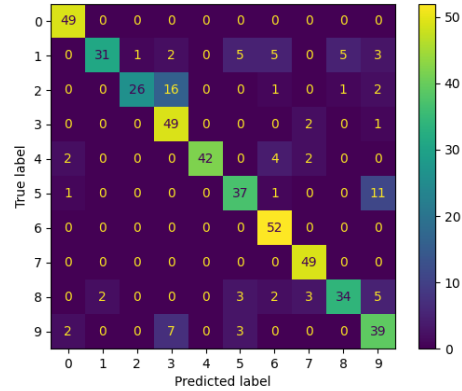
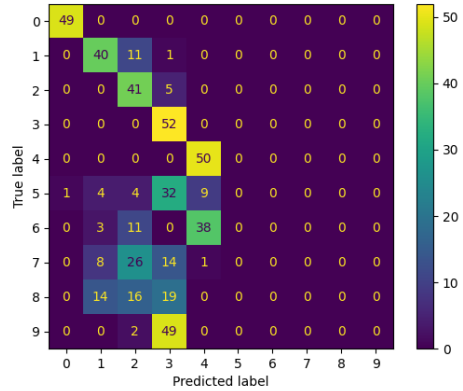
Confusion Matrix Under Dataset With Higher Cluster Diversity  
But Lower Data DiversityConfusion Matrix Under Dataset With Higher Data Diversity  
But Lower Cluster Diversity

Figure 1: Confusion matrix under different training datasets.

## REFERENCES

- [1] Xinyi Xu, Zhaoxuan Wu, Chuan Sheng Foo, and Bryan Kian Hsiang Low. Validation free and replication robust volume-based data valuation. *Advances in Neural Information Processing Systems*, 34, 2021.