A PROOFS AND DERIVATIONS

PROPOSITION 1. ***(Volume wrt. Bias for*** *rank* $= 1$.*) For a matrix* $X \in R^{N\times 1}$, *and its non-zero row submatrices* $X1, X2 \in R^{\frac{N}{2}\times 1}$ *or a matrix* $X \in R^{N\times 2}$ *and its column submatrices* $X1, X2 \in R^{N\times 1}$, *if* $Vol(X_1) \geq Vol(X_2)$, *then* $bias_1 \leq bias_2$.

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,* $\left(\mathbf{X_S}^\top \mathbf{X_S}\right)^{-1} = \|\mathbf{X}_S\|^2 = Vol(X_S)^2$. *Considering the pseudo-inverse is denoted as:* $\mathbf{X_S}^+ := \left(\mathbf{X_S}^\top \mathbf{X_S}\right)^{-1} \mathbf{X_S}^\top$, *the bias variation is:*

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

Since the $Vol(X)$ is constant under $column = 1$, a larger volume yields a smaller bias is proved. Combined with the relation $variation_1 - variation_2 =$ bias $_2 -$ bias $_1$, a larger volume yields a larger parameter variation is apparent.

PROPOSITION 2. ***(Volume wrt. Bias for*** *rank* > 1.*) For a matrix* $X \in R^{N\times M}$, *and its non-zero row submatrices* $X1, X2 \in R^{\frac{N}{2}\times M}$ *or column submatrices* $X1, X2 \in R^{N\times \frac{M}{2}}$, *the following relation can be constructed by Sylvester's formula :*

$$
\begin{aligned}
& bias_2^2 - bias_1^2 = \left\|X^+ - X_2^+\right\|^2 - \left\|X^+ - X_1^+\right\|^2 \\
& = \frac{\left\|Q_2 X_2^T\right\|^2}{Vol(X_2)^4} - \frac{\left\|Q_1 X_1^T\right\|^2}{Vol(X_1)^4} + 2\left\langle \frac{Q_X X^T}{Vol(X)^2}, \frac{Q_1 X_1^T}{Vol(X_1)^2} - \frac{Q_2 X_2^T}{Vol(X_2)^2} \right\rangle,
\end{aligned}
\tag{2}
$$

where,

$$
Q = \textstyle\sum_{l=1}^{k} (\lambda_l \sigma_l)^{-1} \prod_{j=1, j\neq l}^{k} (G - \lambda_i I)\,,
\tag{3}
$$

$$
\sigma_l = \sum_{g=1}^{k} (-1)^{g+1} \lambda_l^{k-g} \left(\textstyle\sum_{\mathcal{H}\subseteq\{1,\ldots,k\}\setminus\{l\}, |\mathcal{H}|=g-1} \left(\prod_{h\in\{1,\ldots,k\}\setminus\mathcal{H}} \lambda_h^{-1} \right) \right).
\tag{4}
$$

$\{\lambda_l, l = 1, \ldots, k\}$ *denotes the left Gram matrix* G*'s* k *unique eigenvalues.*

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

$$
\begin{aligned}
& \mathrm{bias}_2^2 - \mathrm{bias}_1^2 := \left\|\mathbf{X}^+ - \mathbf{X}_2^+\right\|^2 - \left\|\mathbf{X}^+ - \mathbf{X}_1^+\right\|^2 \\
& = \left\|\mathbf{X}^+\right\|^2 - 2\left\langle \mathbf{X}^+, \mathbf{X}_2^+ \right\rangle + \left\|\mathbf{X}_2^+\right\|^2 - \left\|\mathbf{X}^+\right\|^2 + 2\left\langle \mathbf{X}^+, \mathbf{X}_1^+ \right\rangle - \left\|\mathbf{X}_1^+\right\|^2 \\
& = \left\|\mathbf{X}_2^+\right\|^2 - \left\|\mathbf{X}_1^+\right\|^2 + 2\left\langle \mathbf{X}^+, \mathbf{X}_1^+ - \mathbf{X}_2^+ \right\rangle \\
& = \left\|\mathbf{G}_2^{-1}\mathbf{X}_2^\top\right\|^2 - \left\|\mathbf{G}_1^{-1}\mathbf{X}_1^\top\right\|^2 + 2\left\langle \mathbf{G}^{-1}\mathbf{X}^\top, \mathbf{G}_1^{-1}\mathbf{X}_1^\top - \mathbf{G}_2^{-1}\mathbf{X}_2^\top \right\rangle
\end{aligned}
\tag{5}
$$

$\mathbf{G}^{-1}$ *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,*

$$
f(\mathbf{X}) = \sum_{l=1}^{k} f\left(\lambda_l\right) \mathbf{X}_l,
\tag{6}
$$

where λ_l *is the i-th distinct eigenvalue of* X *and* X_l *is the Frobenius covariant as follows:*

$$
\mathbf{X}_l := \prod_{j=1, j\neq l}^{k} \frac{1}{\lambda_l - \lambda_j} \left(\mathbf{X} - \lambda_j \mathbf{I}\right).
\tag{7}
$$

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

$$
f(\mathbf{X}) = \mathbf{X}^{-1} = \sum_{l=1}^{k} \frac{1}{\lambda_l} \mathbf{X}_l
\tag{8}
$$

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

$$\mathbf{G}^{-1} = \sum_{l=1}^{k} \frac{1}{\lambda_l} \mathbf{G}_l = \prod_{j=1, j\neq l}^{k} \frac{1}{\lambda_l - \lambda_j} \prod_{j=1, j\neq l}^{k} \left(\mathbf{G} - \lambda_j \mathbf{I}\right) \tag{9}$$

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

$$\frac{1}{p_l} = \Lambda \underbrace{\sum_{g=1}^{k} (-1)^{g+1} \lambda_l^{k-g} \left[\sum_{\mathcal{H} \subseteq \{1,\ldots,k\} \setminus \{l\}, |\mathcal{H}| = g-1} \left(\prod_{h \in \{1,\ldots,k\} \setminus \mathcal{H}} \frac{1}{\lambda_h} \right) \right]}_{\sigma_l}, \tag{10}$$

where $\Lambda = \prod_{i=l}^{k} \lambda_l = |\mathbf{G}|$. *The Equation 9 can be rewritten as:*

$$\begin{aligned} \mathbf{G}^{-1} &= \frac{1}{|\mathbf{G}|} \underbrace{\frac{1}{\sigma_l} \prod_{j=1, j\neq l}^{k} \left(\mathbf{G} - \lambda_j \mathbf{I}\right)}_{\mathbf{Q}} \\ &= \frac{1}{\mathrm{Vol}(\mathbf{X})^2} \underbrace{\frac{1}{\sigma_l} \prod_{j=1, j\neq l}^{k} \left(\mathbf{G} - \lambda_j \mathbf{I}\right)}_{\mathbf{Q}} \end{aligned} \tag{11}$$

Substituting $\mathbf{G}^{-1}(\mathbf{G_1}^{-1}, \mathbf{G_2}^{-1})$ *in Equation 5 by above result, the Equation 5 can be rewritten as:*

$$\mathrm{bias}_2^2 - \mathrm{bias}_1^2 = \frac{\left\|Q_2 X_2^T\right\|^2}{Vol(X_2)^4} - \frac{\left\|Q_1 X_1^T\right\|^2}{Vol(X_1)^4} + 2\left\langle \frac{Q_X X^T}{Vol(X)^2}, \frac{Q_1 X_1^T}{Vol(X_1)^2} - \frac{Q_2 X_2^T}{Vol(X_2)^2} \right\rangle \tag{12}$$

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

Lemma 2. *For a invertible matrix* A *and any column vector* u *and* v, *then there exits:*

$$\det\left(A + uv^T\right) = \det(A)\left(1 + v^T A^{-1} u\right) \tag{13}$$

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

$$\begin{aligned} Vol(\bar{X})^2 &= \left|\bar{X}^T \bar{X}\right| = \left|\left[X^T, x^T\right] \begin{bmatrix} X \\ x \end{bmatrix}\right| \\ &= \left|X^T X + x^T x\right| = \sigma_1 \left|X^T X\right|. \end{aligned} \tag{14}$$

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* $X' = \left[x_1^T, \ldots, x_n^T\right]^T$ *and adding its row vectors in matrix* X *row by row, the square volume changing is written as:*

$$\begin{aligned} Vol(\left[X^T, X'^T\right]^T) &= Vol(\left[X^T, x_1^T, \ldots, x_n^T\right]^T) \\ &= \sigma_n Vol(\left[X^T, x_1^T, \ldots, x_{n-1}^T\right]^T) = \ldots = \prod_{i=1}^{n} \sigma_i Vol(X), \end{aligned} \tag{15}$$

where $\sigma_i = 1 + x_i(X^T X)^{-1} x_i^T$. *For each* σ_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.*

Proposition 3. ***(Volume is not robust to replication)*** *For a* N*-by-*M *matrix* X, *a replicated* $(N + |X_S|\, d)$*-by-*M *matrix* $X_{rep} := \left[X^T, X_S^T, \ldots, X_S^T\right]^T$ *or a* N*-by-*$(M + |X_S|\, d)$ *matrix* $X_{rep} := [X, X_S, \ldots, X_S]$ *is generated when replicating a row/column submatrix* X_s *in* X *for* $d > 0$ *times. Volume is not robust to replication for* $Vol\left(X_{rep}\right) > Vol\left(X\right)$ *and* $\lim_{d\rightarrow\infty} Vol\left(X_{rep}\right) = \infty$. .

Proof 4. ***(Proposition 3)*** *Considering a replication-involving dataset* $X_{rep} = \left[X^T, \underbrace{X_S^T, ..., X_S^T}_{d} \right]^T$ *, where X_S is row vector and is replicated for d times. According to Equation 15, the square volume of X_{rep} is written as:*

$$Vol(X_{rep})^2 = (1 + X_S(X^T X)^{-1} X_S^T)^d \left| X^T X \right| \tag{16}$$

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

Proposition 4. *$ClusterRV(X)$ is with outlier robustness, for the datasize $N \longrightarrow 0$,*

$$ClusterRV(\{X, x_{outlier}\}) = ClusterRV(X) \tag{17}$$

The proof of Proposition 4 relies on the mechanism that a singleton cluster will be merged to another nearest non-singleton cluster in k-Means. When the dataset size $N \longrightarrow \infty$, the effect of sporadic outliers on cluster center is tending to 0.

Proposition 5. *For $\alpha \in (0, 1)$, the clusterRV's inflation has the following inequality: $(1 - \alpha)^K \le inflation \le (1 - \alpha)^{-K}$, where K is the number of clusters.*

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

$$inflation = \frac{ClusterRV(X_{rep})}{ClusterRV(X)} = \frac{Vol(\widetilde{X}_{rep}) \prod_{i\in K} \rho_{rep,i}}{Vol(\widetilde{X}) \prod_{i\in K} \rho_i} = \frac{\prod_{i\in K} \rho_{rep,i}}{\prod_{i\in K} \rho_i} \tag{18}$$

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

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

$$1 \le \rho_{rep,i} := \textstyle\sum_{p=0}^{\phi_{rep,i}} \alpha^p = \frac{1 - \alpha^{(\phi_{rep,i}+1)}}{1 - \alpha} \le \frac{1}{1 - \alpha} \tag{19}$$

$$1 \le \rho_i := \textstyle\sum_{p=0}^{\phi_i} \alpha^p = \frac{1 - \alpha^{(\phi_i+1)}}{1 - \alpha} \le \frac{1}{1 - \alpha} \tag{20}$$

Thus,

$$(1 - \alpha)^K \le inflation \le (1 - \alpha)^{-K} \tag{21}$$

Proposition 6. *Let $\alpha = 1/(\beta N)$, where N is the size of dataset. If $N \longrightarrow \infty$, for any cluster number K, the inflation of $ClusterRV$ will converge to 1.*

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

$$1 \le \rho_{rep,i} = \textstyle\sum_{p=0}^{\phi_{rep,i}} \frac{1}{\beta N}^p \le \frac{1}{1 - \frac{1}{\beta N}} = 1 + \frac{1}{\beta N - 1} \tag{22}$$

$$1 \le \rho_i = \textstyle\sum_{p=0}^{\phi_i} \frac{1}{\beta N}^p \le \frac{1}{1 - \frac{1}{\beta N}} = 1 + \frac{1}{\beta N - 1} \tag{23}$$

Combined with Equation 18,

$$(1 + \frac{1}{\beta N - 1})^{-K} \le inflation \le (1 + \frac{1}{\beta N - 1})^K \tag{24}$$

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

$$\lim_{N\to\infty} (1 + \frac{1}{\beta N - 1})^{-K} = \lim_{N\to\infty} (1 + \frac{1}{\beta N - 1})^K = 1 \tag{25}$$

Thus, $inflation \longrightarrow 1$ is hold under $N \longrightarrow \infty$.

Proposition 7. *Considering $\alpha = 1/(\beta N)$ and $\beta > 0$. For any fixed cluster number K, if the size of dataset $N \longrightarrow \infty$, the bounded distortion of $ClusterRV$ will converge to 1 .*

Proof 7. ***(Proposition 7)*** *When $N \longrightarrow \infty$, combined with Proposition 6 has proved, the bounded distortion can be rewrittend as:*

$$bounded\_distortion = \frac{ClusterRV(replicate(X_1, c))Vol(X_2)}{Vol(X_1)ClusterRV(replicate(X_2, c))} \approx \frac{Vol(\widetilde{X_1})Vol(X_2)}{Vol(\widetilde{X_2})Vol(X_1)} \tag{26}$$

Taking the scenario where the dataset $X_1 \in R^{N\times M}$ is clustered to K clusters as an example. The matrix of cluster centers is denoted as $\widetilde{X_1} \in R^{K\times M}$. When $N \rightarrow \infty$, the number of data points within each cluster are also infinite. Denoting each clusters $C_i, i = 1, ..., K$ contains D data points evenly, the volume of X_1 is as follows.

$$Vol(X_1)^2 = \left| \left[(\widetilde{X_1} + \Gamma_1)^T, ..., (\widetilde{X_1} + \Gamma_D)^T \right] \begin{bmatrix} \widetilde{X_1} + \Gamma_1 \\ \vdots \\ \widetilde{X_1} + \Gamma_D \end{bmatrix} \right| = \left| \sum_{i=1}^{D} (\widetilde{X_1} + \Gamma_i)^T (X_1 + \Gamma_i) \right| \tag{27}$$

where $\Gamma_i = \begin{bmatrix} C_{0,i} - \widetilde{X}_0 \\ \vdots \\ C_{K,i} - \widetilde{X}_K \end{bmatrix} \in R^{K\times M}$ *denotes the Euclidean distance of i-th data points in each clusters to its cluster's center. Each cluster center generated by k-Means is:*

$$\widetilde{X}_k = \frac{\sum_{i=1}^{D} \delta_{ik} x_i}{\sum_{i=1}^{D} \delta_{ik}}. \tag{28}$$

where δ_{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,

$$\sum_{i=0}^{D} \Gamma_D \rightarrow \overrightarrow{0}. \tag{29}$$

In the same way,

$$Vol(X_2)^2 = \left| \sum_{i=1}^{D} (\widetilde{X}_2)^T (\widetilde{X}_2) + \sum_{i=1}^{D} (\widetilde{\Upsilon}_i)^T (\Upsilon_i) \right| \tag{30}$$

Applying determinant property,

$$\frac{Vol(X_1)^2}{Vol(X_2)^2} = \frac{\left| D\cdot(\widetilde{X}_1)^T(\widetilde{X}_1) + \sum_{i=1}^{D}(\Gamma_i)^T(\Gamma_i) \right|}{\left| D\cdot(\widetilde{X}_2)^T(\widetilde{X}_2) + \sum_{i=1}^{D}(\Upsilon_i)^T(\Upsilon_i) \right|} = \frac{\left| (\widetilde{X}_1)^T(\widetilde{X}_1) + \frac{1}{D}\sum_{i=1}^{D}(\Gamma_i)^T(\Gamma_i) \right|}{\left| (\widetilde{X}_2)^T(\widetilde{X}_2) + \frac{1}{D}\sum_{i=1}^{D}(\Upsilon_i)^T(\Upsilon_i) \right|} \tag{31}$$

Considering $\Gamma_i^T \Gamma_i = \alpha_i \widetilde{X}_1^T \widetilde{X}_1$, for the data size $N \rightarrow \infty$, there will also exists Υ_i in infinite space satisfied $\Upsilon_i^T \Upsilon_i = \alpha_i \widetilde{X}_2^T \widetilde{X}_2$. Whether $\Gamma_{i,k}$ and $\Upsilon_{i,k}$ belong to same clusters k is not necessary, for:

$$\Gamma_i^T \Gamma = \begin{bmatrix} \Gamma_{i,1}^T & \dots & \Gamma_{i,K}^T \end{bmatrix} \begin{bmatrix} \Gamma_{i,1} \\ \dots \\ \Gamma_{i,K} \end{bmatrix} = \sum_{k=1}^{K} \Gamma_{i,k}^T \Gamma_{i,k} = \underbrace{\begin{bmatrix} \Gamma_{i,K}, \dots & \Gamma_{i,1} \end{bmatrix}}_{any\ order} \begin{bmatrix} \Gamma_{i,K} \\ \dots \\ \Gamma_{i,1} \end{bmatrix} \tag{32}$$

Then, the Equation 31 is rewritten as:

$$\frac{Vol(X_1)^2}{Vol(X_2)^2} = \frac{\left| (\widetilde{X}_1)^T(\widetilde{X}_1) + \frac{1}{D}\sum_{i=1}^{D}(\Gamma_i)^T(\Gamma_i) \right|}{\left| (\widetilde{X}_2)^T(\widetilde{X}_2) + \frac{1}{D}\sum_{i=1}^{D}(\Upsilon_i)^T(\Upsilon_i) \right|} = \frac{\left| I + \frac{1}{D}\sum_{i=1}^{D}\alpha_i \right|}{\left| I + \frac{1}{D}\sum_{i=1}^{D}\alpha_i \right|} \frac{\left| (\widetilde{X}_1)^T(\widetilde{X}_1) \right|}{\left| (\widetilde{X}_2)^T(\widetilde{X}_2) \right|} = \frac{Vol(\widetilde{X}_1)^2}{Vol(\widetilde{X}_2)^2} \tag{33}$$

Thus, when $N \rightarrow \infty$, the data points in each cluster $D \rightarrow \infty$,

$$bounded\_distortion \approx \frac{Vol(\widetilde{X_1})Vol(X_2)}{Vol(\widetilde{X_2})Vol(X_1)} \rightarrow 1. \tag{34}$$

Definition 1. ***(Unbounded subset-sum problem)*** *Given a set of positive integers $\{k_0, ..., k_n\}$, an unbounded subset-sum problem is defined as to find the non-negative integers α_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 $v_i = p_i, i = 1, ..., n$, $p_{n+1} = K + \triangle$ when $v_{n+1} = K$ and $\triangle \in (0, 1)$, 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:*

$$K + \triangle = p(K) = p(\sum_{i=1}^{n} \alpha_i v_i) \leq \sum_{i=1}^{n} \alpha_i p_i \overset{v_i = p_i}{=} K \tag{35}$$

$K + \triangle = K$ 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 contrains an unbounded subset sum constains* x*, thus* $(x) \geq x$*. Then, we define a function* $p(x) = min((x), K + \triangle)$ *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 minumum unbounded subset sum, we have* $_i = v_i = p_i \leq K + \triangle$*. 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:*

$$p(x) = \begin{cases} \mu(x), & \mu(x) \leq K \\ K + \triangle, & \mu(x) \geq K + 1 \end{cases} \tag{36}$$

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

Above all, we have proved Lemma 3). For the unbounded subset-sum problem is NP-hard, proving the unbounded subset sum $\sum_{i=1}^{n} \alpha_i v_i = K$ not exists in Lemma 3) is a co-NP hard problem.

Proposition 8. *Given a pricing function p satisfying the constraint of* $p(x)/x \geq p(y)/y, x \leq y$*, it also satisfy* $p(x + y) \leq p(x) + p(y)$ *strictly.*

Proof 9. ***(Proposition 8)*** *For the pricing function p satisfies* $p(x)/x \geq p(y)/y$ *when* $x \leq y$*, there exists:*

$$\begin{aligned} \frac{p(x+y)}{x+y} &\leq \min\left(\frac{p(x)}{x}, \frac{p(y)}{y}\right) \Rightarrow \\ p(x+y) &\leq \min\left(p(x) + \underbrace{\frac{yp(x)}{x}}_{\geq p(y)}, p(y) + \underbrace{\frac{xp(y)}{y}}_{\geq p(x)}\right) \\ &\leq p(x) + p(y) \end{aligned} \tag{37}$$

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 a 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.*

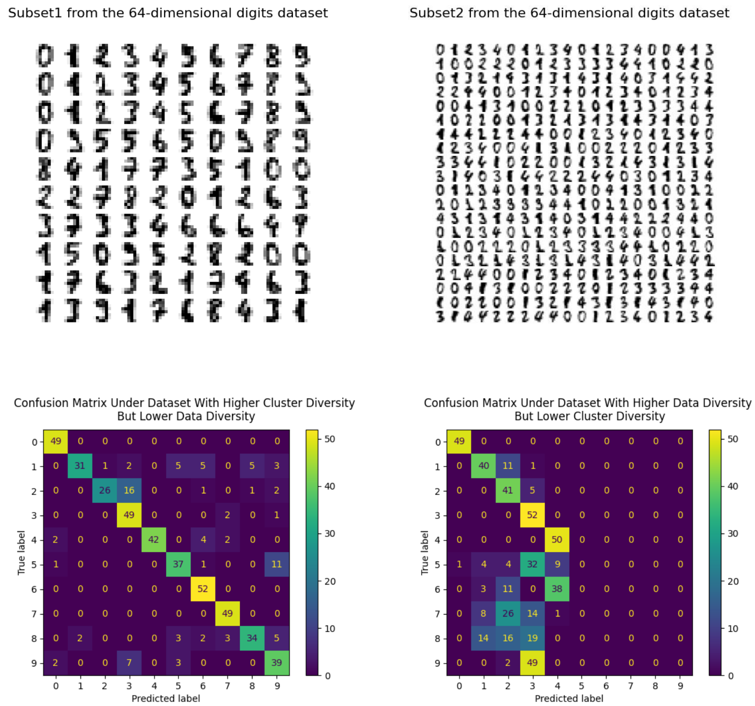

Figure 1: Confusion matrix under different training datasets.