Towards Discriminability and Diversity: Batch Nuclear-norm Maximization under Label Insufficient Situations

A. Appendix

A.1. Monotonicity of F-norm and Entropy

To prove the strict opposite monotonicity between F-norm and entropy, we seperately analyze their monotonicity and then compare their bounds and monotonicity.

The F-norm of matrix A is the square root sum of all the elements in A. The calculation process could be divided into two process, we could first calculate the quadratic sum of each row in A and then calculated the square root of sum of all the rows. Besides the condition that the monotonicity of square root of sum of all the rows depends on the monotonicity of each row, there is no extra constraint on different rows. Thus we could simply consider the monotonicity of quadratic sum of each row to analyze the monotonicity of the F-norm on matrix. Similar for the entropy, we could also simply analyze the monotonicity of each row.

We take the row i for example, and denote the square sum of row i as $f(\mathrm{A}_i)$, thus $f(\mathrm{A}_i)$ could be calculated as:

$$f(\mathrm{A}_i) = \sum_{j=1}^{C} \mathrm{A}_{ij}^2. \tag{1}$$

To analyze the monotonicity of a function of several input variables, we could analyze the monotonicity of the function on each variable. It should be noted that $\sum_{j=1}^{C} \mathrm{A}_{ij} = 1$. Actually the variables are supposed to be independent, but to satisfy the constraint of the sum 1, we assume the variable A_{iC} is the only variable dependent on A_{ij}. Thus the partial derivative of $f(\mathrm{A}_i)$ could be calculated as:

$$\begin{aligned}\frac{\partial f(\mathrm{A}_i)}{\partial \mathrm{A}_{ij}} &= 2\mathrm{A}_{ij} - 2\mathrm{A}_{iC} \\ &= 4\mathrm{A}_{ij} - 2(1 - \sum_{k=1,k\neq j}^{C-1} \mathrm{A}_{ik}),\end{aligned} \tag{2}$$

where $f(\mathrm{A}_i)$ reachs the bound when $\mathrm{A}_{ij} = \frac{1}{2} - \frac{1}{2}\sum_{k=1,k\neq j}^{C-1} \mathrm{A}_{ik}$. When $\mathrm{A}_{ij} \leq \frac{1}{2} - \frac{1}{2}\sum_{k=1,k\neq j}^{C-1} \mathrm{A}_{ik}$ and $\frac{\partial f(\mathrm{A}_i)}{\partial \mathrm{A}_{ij}} \leq 0$, $f(\mathrm{A}_i)$ will monotonously decrease. When $\mathrm{A}_{ij} \geq \frac{1}{2} - \frac{1}{2}\sum_{k=1,k\neq j}^{C-1} \mathrm{A}_{ik}$ and $\frac{\partial f(\mathrm{A}_i)}{\partial \mathrm{A}_{ij}} \geq 0$, $f(\mathrm{A}_i)$ will monotonously increase.

For the entropy, we denote the entropy of row i as $h(\mathrm{A}_i)$, and $h(\mathrm{A}_i)$ could be calculated as follows:

$$h(\mathrm{A}_i) = -\sum_{j=1}^{C} \mathrm{A}_{ij} \log(\mathrm{A}_{ij}). \tag{3}$$

Similarly, the partial derivative of $h(\mathrm{A}_i)$ could be calculated as:

$$\begin{aligned}\frac{\partial h(\mathrm{A}_i)}{\partial \mathrm{A}_{ij}} &= -\log(\mathrm{A}_{ij}) + \log(\mathrm{A}_{iC}) \\ &= \log(\frac{1 - \mathrm{A}_{ij} - \sum_{k=1,k\neq j}^{C-1} \mathrm{A}_{ik}}{\mathrm{A}_{ij}})\end{aligned} \tag{4}$$

where $h(\mathrm{A}_i)$ reachs the bound when $\mathrm{A}_{ij} = \frac{1}{2} - \frac{1}{2}\sum_{k=1,k\neq j}^{C-1} \mathrm{A}_{ik}$. When $\mathrm{A}_{ij} \leq \frac{1}{2} - \frac{1}{2}\sum_{k=1,k\neq j}^{C-1} \mathrm{A}_{ik}$ and $\frac{\partial h(\mathrm{A}_i)}{\partial \mathrm{A}_{ij}} \geq 0$, $h(\mathrm{A}_i)$ will monotonously increase. When $\mathrm{A}_{ij} \geq \frac{1}{2} - \frac{1}{2}\sum_{k=1,k\neq j}^{C-1} \mathrm{A}_{ik}$ and $\frac{\partial h(\mathrm{A}_i)}{\partial \mathrm{A}_{ij}} \leq 0$, $h(\mathrm{A}_i)$ will monotonously decrease. This validates that the F-norm and entropy of matrix $f(\mathrm{A}_i)$ have strict opposite monotonicity.

A.2. Relationship between Nuclear-norm and F-norm

We reanalyze the relation between $\|\mathrm{A}\|_*$ and $\|\mathrm{A}\|_F$. We denote $\times$ as the matrix multiplication and the trace of matrix $\mathrm{A} \times \mathrm{A}^\top$ is as follows:

$$\begin{aligned}trace(\mathrm{A} \times \mathrm{A}^\top) &= \sum_{i=1}^{B}\sum_{j=1}^{C} \mathrm{A}_{i,j} \cdot \mathrm{A}_{i,j} \\ &= (\|\mathrm{A}\|_F)^2\end{aligned} \tag{5}$$

The trace of $\mathrm{A} \times \mathrm{A}^\top$ equals to the sum of eigenvalues of $\mathrm{A} \times \mathrm{A}^\top$. While the calculated eigenvalues of $\mathrm{A} \times \mathrm{A}^\top$ is the square of singular value of A. We denote the ith largest singular value as σ_i. Thus $trace(\mathrm{A} \times \mathrm{A}^\top)$ becomes quadratic sum of singular values of matrix A:

$$trace(\mathrm{A} \times \mathrm{A}^\top) = \sum_{i=1}^{D} \sigma_i^2. \tag{6}$$

Combining Eqn. 5 and 6, we could find that:

$$(\|\mathrm{A}\|_F) = \sqrt{\sum_{i=1}^{D} \sigma_i^2}, \tag{7}$$

where the number of the singular values is denoted as D and $D = Min(B, C)$. For the matrix A, the calculation of the nuclear-norm could be achieved by the sum of singular values of A. Thus the nuclear-norm could be calculated as follows:

$$\|\mathrm{A}\|_* = \sum_{i=1}^{D} \sigma_i, \tag{8}$$

Thus we could find the upper-bound of $\|\mathrm{A}\|_*$ as:

$$\|\mathrm{A}\|_* = \sqrt{(\sum_{i=1}^{D} \sigma_i)^2} \leq \sqrt{D \cdot \sum_{i=1}^{D} \sigma_i^2} = \sqrt{D} \cdot \|\mathrm{A}\|_F\,, \tag{9}$$

where if $\|\mathrm{A}\|_* = \sqrt{D} \cdot \|\mathrm{A}\|_F$, all the singular values will be the same. Similarly, we could obtain the lower-bound of $\|\mathrm{A}\|_*$ as:

$$\|\mathrm{A}\|_* = \sqrt{(\sum_{i=1}^{D} \sigma_i)^2} \geq \sqrt{\cdot \sum_{i=1}^{D} \sigma_i^2} = \|\mathrm{A}\|_F\,, \tag{10}$$

Combining Eqn. 9 and 10, we could summarize the relationship as follows:

$$\frac{1}{\sqrt{D}} \|\mathrm{A}\|_* \leq \|\mathrm{A}\|_F \leq \|\mathrm{A}\|_* \leq \sqrt{D} \cdot \|\mathrm{A}\|_F\,. \tag{11}$$

Thus $\|\mathrm{A}\|_*$ and $\|\mathrm{A}\|_F$ could bound each other.

A.3. The Nuclear-norm Calculation in Eqn. (8)

We assume B and C are 2. In this case, A could be expressed as:

$$\mathrm{A} = \begin{bmatrix} x & 1-x \\ y & 1-y \end{bmatrix}, \tag{12}$$

where x and y are variables. To calculate the singular values, we build a new matrix $\mathrm{A} \times \mathrm{A}^\top$ as follows:

$$\mathrm{A} \times \mathrm{A}^\top = \begin{bmatrix} x^2 + (1-x)^2 & xy + (1-x)(1-y) \\ xy + (1-x)(1-y) & y^2 + (1-y)^2 \end{bmatrix}, \tag{13}$$

where we could calculate the eigen values of matrix $\mathrm{A} \times \mathrm{A}^\top$ by:

$$|\mathrm{A} \times \mathrm{A}^\top - \lambda \mathcal{I}| = 0. \tag{14}$$

Thus we could substitute into the value of $\mathrm{A} \times \mathrm{A}^\top$ as follows:

$$\begin{vmatrix} x^2 + (1-x)^2 - \lambda & xy + (1-x)(1-y) \\ xy + (1-x)(1-y) & y^2 + (1-y)^2 - \lambda \end{vmatrix} = 0. \tag{15}$$

By integrating the equation, we could obtain the following results:

$$\lambda^2 - 2(x^2 - x + y^2 - y + 1)\lambda + (y - x)^2 = 0. \tag{16}$$

Considering that there are only two singular values in this situation, we denote them as σ_1 and σ_2. While the solution of the Eqn. 16 is the square of the singular values. Thus we could find that:

$$\begin{aligned} \sigma_1^2 + \sigma_2^2 &= 2(x^2 - x + y^2 - y + 1) \\ \sigma_1^2 \cdot \sigma_2^2 &= (y - x)^2 \end{aligned} \tag{17}$$

The sum of the singular values is calculated as follows:

$$\begin{aligned} \sigma_1 + \sigma_2 &= \sqrt{(\sigma_1 + \sigma_2)^2} \\ &= \sqrt{(\sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2} \\ &= \sqrt{(2(x^2 - x + y^2 - y + 1) + 2|y - x|} \end{aligned} \tag{18}$$

Then the nuclear-norm could be calculated as:

$$\|\mathrm{A}\|_* = \sqrt{x^2 + (1-x)^2 + y^2 + (1-y)^2 + 2|y - x|}, \tag{19}$$