Dimension and Rank of Effective Input Matrix

Minhuan Li   minhuanli@g.harvard.edu

This is a discussion document targeting the dimension and rank of effective input matrix in mixed selectivity project, based on Arvin's thesis. I will give the expressions of those two important parameters, different from that in Arvin's thesis.

1 Basic Framework

This discussion is based on CEMS model in Arvin's thesis. The basic framework is described below:

1.1 Input layer, Mixed layer and Readout layer

3 layers in total, first layer is input layer, second layer is mixed layer, and third layer is linear readout layer.

Input layer gathers information from N_m independent modalities, in which there are $N_m - 1$ contexts and 1 task-relevant stimuli. Dimension of contextual vectors is M and dimension of stimuli vectors is N. So the dimension of general vector in input layer is $N + M(N_m - 1)$. Stimuli modality has P independent samples, while each contextual modality has K independent samples. The total input data matrix is $\xi \in \mathbb{R}^{N+M(N_m-1), PK^{N_m-1}}$

Input layer connects to the second layer in a manner called mixed selectivity. Given the mixing index $\mathcal{M}$, each neuron in second layer non-linearly mixes information of $\mathcal{M} - 1$ contextual modalities and 1 stimuli modality in input layer. The connection is realized by random matrix $\mathbf{J}$.

Then a linear readout based on the second layer gives classification.

1.2 Parameters list

N_m	Modality Number
N	Task-Relevant Stimuli Vector Dimension
P	Number of Independent Stimuli Samples
M	Context Vector Dimension
K	Number of Independent Contextual Samples in each Modality
$\mathcal{M}$	Mixing Index
N_c	Vector Dimension in the Mixed Layer
ϵ	Notation to Represent Stimuli Data
σ, μ, η	Notations to Represent Contextual Data
$\mathcal{P}$	Number of Partitions in Mixing

2 Effective Input Matrix

Effective input matrix is the data matrix of mixed layer. As the linear readout is implemented after mixed layer, properties of effective input matrix are strongly relevant to the classification capacity.

2.1 Why we care about rank r?

According to ShK theory[1], the capacity of correlated dataset is:

$$\alpha_c = 2c \tag{1}$$

$c = r/N$, while N is the dimension of data vector, r is the matrix rank.

2.2 Partition Scheme and Dimension of Effective Input Matrix

The first dim is N_c, the length of vector. The second dim corresponds to the number of data combinations. Given we have N_m modality and mixing index is $\mathcal{M} > 1$. The the number of partitions, groups of stimuli and contextual modalities mixture, is:

$$\mathcal{P}(N_m, \mathcal{M}) = \binom{N_m - 1}{\mathcal{M} - 1} \tag{2}$$

Specially, $\mathcal{P} = N_m$ if $\mathcal{M} = 1$. In each partition, number of independent data combinations is

$$PK^{\mathcal{M}-1}$$

So, the total number of data combinations, as well as the second dim, is:

$$[PK^{\mathcal{M}-1}]^{\mathcal{P}}, \quad \text{if } \mathcal{M} > 1 \tag{3}$$

If $\mathcal{M} = 1$, the second dim is:

$$PK^{N_m-1}, \quad \text{if } \mathcal{M} = 1 \tag{4}$$

So the dimension of effective input matrix is:

$$\bar{\xi} \in \mathbb{R}^{N_c \times [PK^{\mathcal{M}-1}]^{\mathcal{P}}}, \quad \text{if } \mathcal{M} > 1 \tag{5}$$

$$\bar{\xi} \in \mathbb{R}^{N_c \times PK^{N_m-1}}, \quad \text{if } \mathcal{M} = 1 \tag{6}$$

For example, set $N_m = 3$ and $\mathcal{M} = 2$, a partial selectivity one. We have stimuli data $\{\epsilon_1, \dots, \epsilon_P\}$, context data $\{\sigma_1, \dots, \sigma_K\}$ and $\{\eta_1, \dots, \eta_K\}$. Then we have two partition scheme:

$$(\epsilon, \sigma) \quad \text{and} \quad (\epsilon, \eta)$$

[1] Takashi Shinzato and Yoshiyuki Kabashima. "Perceptron capacity revisited: clas- sification ability for correlated patterns". In: Journal of Physics A: Mathematical and Theoretical 41.32 (2008), p. 324013.

The effective data matrix is like:

$$\begin{bmatrix} h(\epsilon_1,\sigma_1), & \dots, & h(\epsilon_P,\sigma_K), & h(\epsilon_1,\sigma_1), & \dots, & h(\epsilon_P,\sigma_K), & \dots, & h(\epsilon_1,\sigma_1), & \dots, & h(\epsilon_P,\sigma_K) \\ h(\epsilon_1,\eta_1), & \dots, & h(\epsilon_1,\eta_1), & h(\epsilon_1,\eta_2), & \dots, & h(\epsilon_1,\eta_2), & \dots, & h(\epsilon_P,\eta_K), & \dots, & h(\epsilon_P,\eta_K) \end{bmatrix} \tag{7}$$

The row number is N_c and the column number is P^2K^2, which agree with the general expression we got above. For $N_m = 3$ and $\mathcal{M} = 1$, dimension is $N_c \times PK^2$; $N_m = 3$ and $\mathcal{M} = 3$, dimension is again $N_c \times PK^2$

2.3 Rank of Effective Input Matrix

Again set $N_m = 3$ and $\mathcal{M} = 2$, a partial selectivity one, as example. The effective data matrix is:

$$\begin{bmatrix} h(\epsilon_1,\sigma_1), & \dots, & h(\epsilon_P,\sigma_K), & h(\epsilon_1,\sigma_1), & \dots, & h(\epsilon_P,\sigma_K), & \dots, & h(\epsilon_1,\sigma_1), & \dots, & h(\epsilon_P,\sigma_K) \\ h(\epsilon_1,\eta_1), & \dots, & h(\epsilon_1,\eta_1), & h(\epsilon_1,\eta_2), & \dots, & h(\epsilon_1,\eta_2), & \dots, & h(\epsilon_P,\eta_K), & \dots, & h(\epsilon_P,\eta_K) \end{bmatrix} \tag{8}$$

Make column subtractions to transfer the matrix. We can view every PK columns as a block (PK blocks in total). In each block, subtract the first column from the remaining columns, we get:

$$\begin{bmatrix} h(\epsilon_1,\sigma_1), & h(\epsilon_1,\sigma_2)-h(\epsilon_1,\sigma_1), & \dots, & h(\epsilon_P,\sigma_K)-h(\epsilon_1,\sigma_1), & \dots, & h(\epsilon_1,\sigma_1), & h(\epsilon_1,\sigma_2)-h(\epsilon_1,\sigma_1), & \dots, & h(\epsilon_P,\sigma_K)-h(\epsilon_1,\sigma_1) \\ h(\epsilon_1,\eta_1), & 0, & \dots, & 0, & \dots, & h(\epsilon_P,\eta_K), & 0, & \dots, & 0 \end{bmatrix} \tag{9}$$

Then we could easily find redundant columns: except the first columns, all other columns are identical between different blocks. Then, make further subtractions to make it clear:

$$\begin{bmatrix} h(\epsilon_1,\sigma_1), & h(\epsilon_1,\sigma_2)-h(\epsilon_1,\sigma_1), & \dots, & h(\epsilon_P,\sigma_K)-h(\epsilon_1,\sigma_1), & ,h(\epsilon_1,\sigma_1), & 0, & \dots, & 0, & \dots, & h(\epsilon_1,\sigma_1), & 0, & \dots, & 0 \\ h(\epsilon_1,\eta_1), & 0, & \dots, & 0, & h(\epsilon_1,\eta_2), & 0, & \dots, & 0, & \dots, & h(\epsilon_P,\eta_K), & 0, & \dots, & 0 \end{bmatrix} \tag{10}$$

So the first block has PK nonzero columns, while each the remaining blocks only have one nonzero column, so the column rank is:

$$PK + PK - 1 = 2PK - 1 \tag{11}$$

Or it can be computed from reducing redundant columns: $P^2K^2 - (PK-1)(PK-1) = 2PK - 1$

So the rank of $N_m = 3$ and $\mathcal{M} = 2$ effective input matrix is:

$$\min\left(N_c,\ 2PK - 1\right) \tag{12}$$

In general, for a mixed-selectivity system with N_m and $\mathcal{M} > 1$, the rank of effective input matrix is:

$$\min\left(N_c,\ PK^{\mathcal{M}-1} + [PK^{\mathcal{M}-1}]^{\mathcal{P}-1} - 1\right), \quad \text{if } \mathcal{M} > 1 \tag{13}$$

if $\mathcal{M} = 1$, the rank is:

$$\min\left(N_c,\ P + K^{N_m-1} - 1\right), \quad \text{if } \mathcal{M} = 1 \tag{14}$$