Preparation

For each sequence (or data point), the image will get a value based on which frame it is in. The formula is:

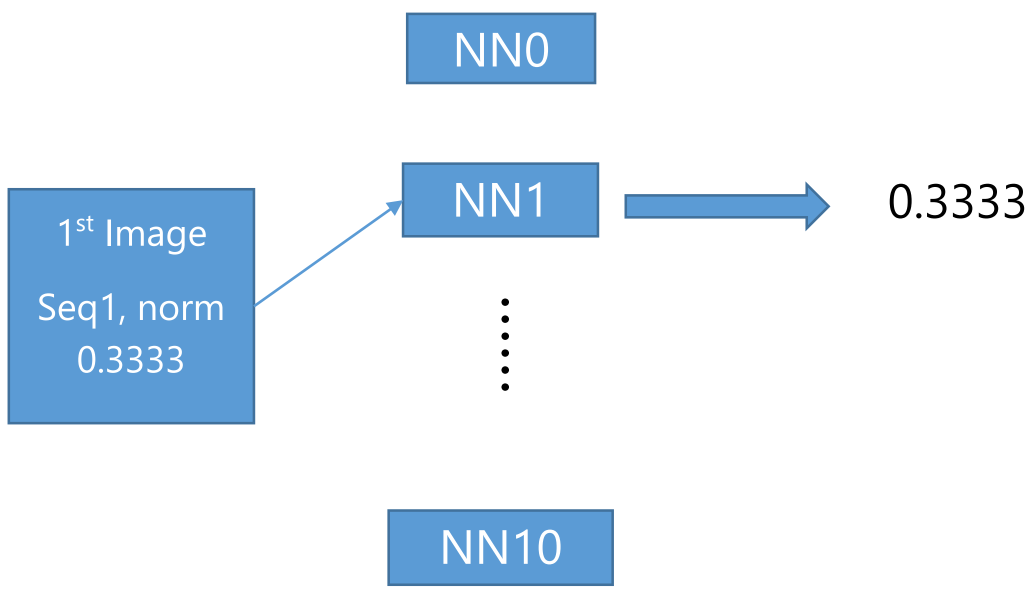
$a_{ij} = \frac{s_{ij}}{n_{j-1}}$, where n_j is the number of images in sequence i, and s_{ij} is the sequential number the image has (sequential number start at 0).

For e.g., there is a frame in training sequence 0, and it is the 5th image in that sequence 0, and that sequence has 13 images, the a_{ij} value could be

$$\frac{5-1}{13-1} = 0.333$$

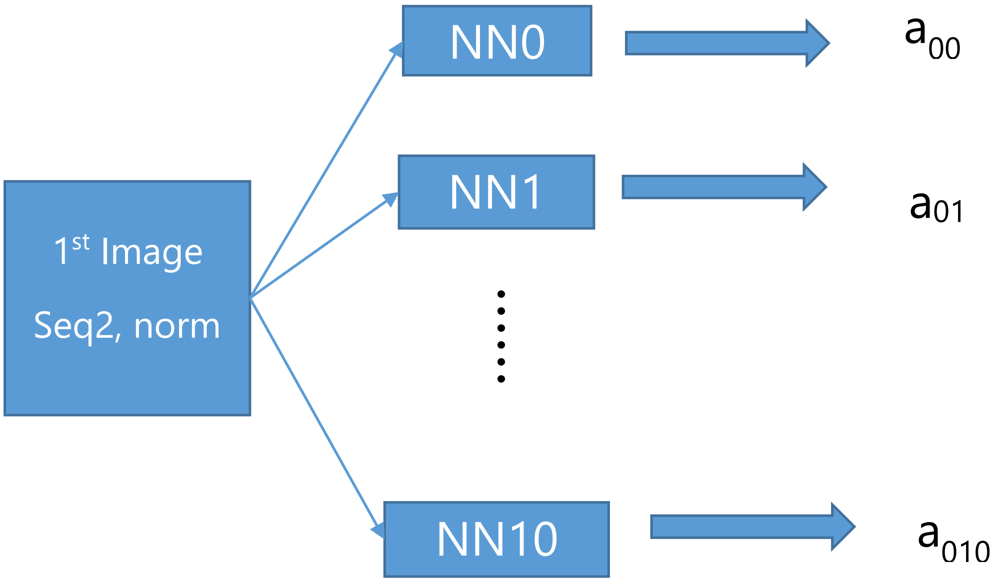
Training

There will be 11 neuron networks, one for each gesture. However, these are only called "neuron networks" at the moment. We will need to investigate what function best suits this situation.



Testing

During testing, each image in a test sequence will be passed through all the neurons networks.



Then for the entire sequence, we will have a matrix, assuming there are k gestures, and j images in a sequence, and this is the ith sequence in the test set.

$$\boldsymbol{S}_i = \begin{bmatrix} a_{00} & a_{01} & \cdots & a_{0k} \\ a_{10} & a_{11} & \cdots & a_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{j0} & a_{j1} & \cdots & a_{jk} \end{bmatrix}$$

In this matrix, each row is data from one image. Therefore, the most evenly distributed columns will be the best fit. To measure "the most evenly distributed column", we can use MSE. To compute this MSE, we need to generate a standard benchmark matrix for k gestures.

$$\boldsymbol{B} = \begin{bmatrix} \frac{0}{k-1} & \frac{1}{k-1} & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{0}{k-1} & \frac{1}{k-1} & \cdots & 1 \end{bmatrix}^T$$

Multiple this with each column of S_i, we can get the MSE for each column. The MSE matrix would be:

$$\boldsymbol{M} = \boldsymbol{B} \times \boldsymbol{S_i}$$

The column with min value in M would be the best gesture.

More discussion can be had. For e.g., what if all MSEs are terrible or good? That would be left to another discussion