Theoretical Neuroscience II
Teacher: Prof. Jochen Braun
Assistant Teacher: Ehsan Kakaei
Name: Viviana Villafaña Barraza & Tasnuba Tazeen & Lee Po Shing
Date: 4th May, 2018

Exercise 3: Supervised Learning

Introduction

Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs. In the process of supervised learning, neurons in the upstream population firstly receive the same input as neurons of the downstream population, and a connection matrix is constructed based on their correlation. If the network is functioning, this will eventually allow the downstream population to respond the same way regardless the input is directly provided to them or indirectly via the upstream population, as a proper connection matrix will help matching all the pairs with similar preferences.

In this assignment, the activity of a network consisting of a unidirectional connection from one population of neurons to another was investigated. In particular, the network was first trained to obtain the connection matrix using same stimulation to both populations separately. After the training, the network was tested to justify the wellness of the connection, i.e. if the information flowed correctly from upstream to downstream population, by stimulating the downstream indirectly with the upstream population. Lastly, by varying the turning curves and the sizes of the two populations, the effect on the connection matrix was analysed.

Method

In this exercise, a network constructed with an upstream populations *N* and a downstream population *M* was trained so that the unidirectional connection from *N* to *M* allows *M* to response as their pre-set turning curves. It was then tested with different parameters to observe the effect on the testing results. The values of the following variables remained unchanged unless specified: the numbers of neurons for *N* and *M* *i* and *j* (also in the code), 50 and 100 respectively; a variable for the deviation of the turning curves of *N* κ (kappa in the code), 0.22; a variable for the deviation of the turning curves of *M* σ (sigma in the code), 30; the maximum response firing rate of neuron R_{max} (R_max in the code), 1 *Hz*; the maximum sound frequency as input v_{max} (nu_max in the code), 1000 *Hz*.

In the training session, the preferred frequencies for *N* and *M* $v_{pref,\,i}$ and $v_{pref,\,j}$ (nu_pref_i and nu_pref_j in the code) were calculated randomly by selecting a value between 0 and v_{max} for each neuron. The responses of *N* R_i (R_i in the code) are based on a Gaussian distribution around the preferred frequencies with a deviation linear to the preferred frequencies; whereas of *M* R_j (R_j in the code) with a constant deviation. The two were calculated as followed:

$$R_i = R_{max}e^{[-\frac{(\nu_{pref,i}-\nu)^2}{2\kappa^2\nu_{pref,i}{}^2}]} \qquad\qquad R_j = R_{max}e^{[-\frac{(\nu_{pref,j}-\nu)^2}{2\sigma^2}]}$$

A random noise was then added based on a normal distribution around 0 with the deviation of the calculated R_i and R_j. The resulting responses were $R_{noisy,\,i}$ and $R_{noisy,\,j}$ (R_noisy_i and R_noisy_j in the code), and they were corrected to 0 if the value is negative. With a vector of training sound frequencies $v_{training}$ (nu_training in the code) of training trails of 1000 (trial_training, in the code), R_i and $R_{noisy,\,i}$ were calculated using the provided Matlab function *GaussResp_LinearSTD.m*; and $R_{noisy,\,j}$ and R_j *GaussResp_ConstantSTD.m*. Below showed the responses of three chosen neurons to $v_{training}$ for both populations.

Next the covariance matrices connecting the two populations were calculated using the provided Matlab function *covariance.m,* and the matrices were plotted without noise W_{ji} and with noise $W_{noisy,\,ji}$ (W and W_noisy in the code) using the Matlab function *pcolor*. The equation for both covariance matrices was as followed:

$$W_{ji} = \overline{\sum_{k=1}^{trial\_training} R_\iota R_\jmath} - \bar{R}_\jmath{}^T \bar{R}_\iota$$

To test the feed-forward connection, $R_{test,\,j}$ (R_test_j in the code) were used to calculate an inferred value of testing sound frequency $v_{inferred}$ (nu_inf in the code) and compared with the actual testing sound frequency v_{test} (nu_test in the code). 1000 trials (trial_test in the code) were used and v_{test} was calculated based on it. Using the provided Matlab function *GaussResp_LinearSTD.m* again, $R_{test,\,noisy,\,i}$ (R_test_noisy_i in the code) were calculated and thus $R_{test,\,j}$ based on $R_{test,\,noisy,\,i}$ and $W_{noisy,\,ji}$ as below:

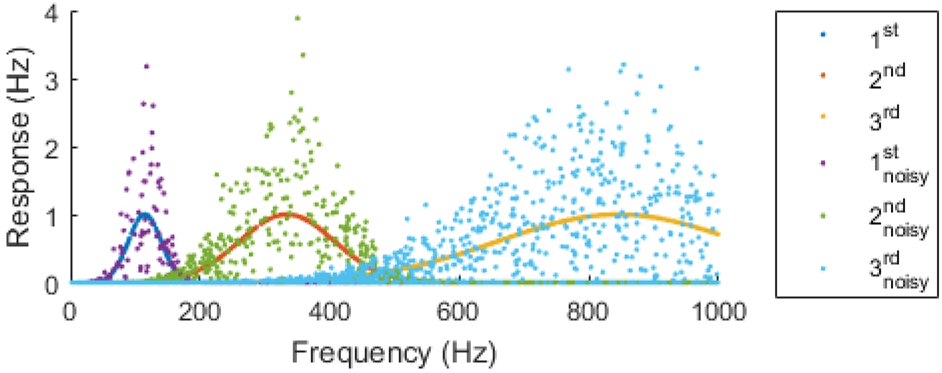
$$R_{test,j} = R_{test,noisy,i}W_{noisy,ji}$$

Any negative value was then corrected to 0. The testing frequencies were inferred as followed:

$$\nu_{inferred} = \frac{R_{test,j}\nu_{pref,j}{}^T}{\sum_{k=1}^{trial\_test} R_{test,j}}$$

The actual testing frequency against the inferred frequency then was plotted. Lastly, the values of κ, σ, *i* and *j* were varied to examine their effects on the training result.

(a)



(b)

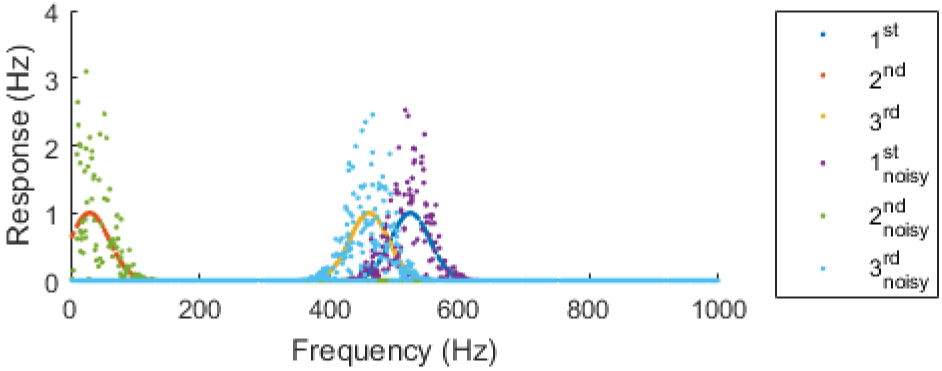


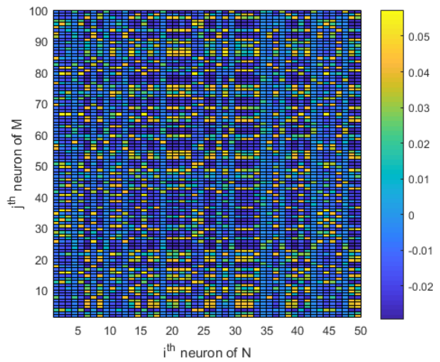
Fig. 2 Representative responses to different stimulating frequencies from upstream and downstream population without or with noise. In the response sets without noise (solid lines), both the upstream (a) and the downstream neurons (b) showed a Gaussian distribution with a linear and constant deviation respectively. For the responses sets with noise (scattered dots), the same upstream and downstream neurons showed responses with deviation from the one with noise.

Results & Discussion

Covariance matrices were successfully obtained after training

As in *Fig. 3*, the two covariance matrices W_{ji} and $W_{noisy,\, ji}$ were computed. Considering the cluster in the middle bottom of *Fig. 3* (a), the preferred frequencies of 30th – 32nd of *N* were 881, 921 and 619 *Hz* while those of 3rd – 7th of *M* were 339, 844, 831, 772 and 193 *Hz*. Therefore, those of *N* should show higher covariance with 4th – 6th of *M* while lower covariance with 3rd and 7th. This has been successfully reflected in the covariance matrices by warmer and colder colours in the respective cells.

(a)



(b)

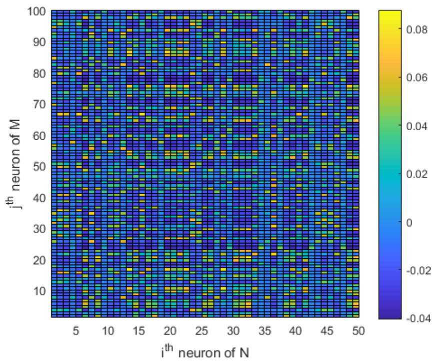


Fig. 3 Representative covariance matrices without and with noise. The warmer colors represented the higher covariance of the neuron pairs between upstream and downstream populations; whereas the colder ones represented the lower covariance. Noted that the scale of the covariance of generated from the noisy responses $W_{noisy,\, ji}$ (b) is wider than that from the ones without noise W_{ji} (a).

Inferred stimulating frequencies closely approximated the actual values

Using the covariance matrix calculated above, the downstream neurons were stimulated indirectly using the response from the upstream population to obtain the $R_{test,\ j}$ and hence $v_{inferred}$. By comparing $v_{inferred}$ with v_{test}, one can estimate the performance of the connection matrix, as a perfect connection matrix should be able to direct the correct information to the right target. Metaphorically saying, the stimulus is just like voice that being recorded as electric signals ($R_{test,\ i}$), and when one wants to play it out again as sound track ($R_{test,\ j}$), a software is needed to convert the signals back to the sound, which is the $W_{noisy,\ ji}$. Therefore, if the sound track does not resemble the original sound, it means that the software is problematic. In our case, it means the feed-forward connection didn't train well.

In *Fig. 4*, $v_{inferred}$ had similar values as v_{test} in most frequencies, except high frequency region of > 800 *Hz*. This can be due to the fact that the deviation of $R_{test,\ i}$ had a linear relationship with $v_{pref,\ i}$, therefore the connection matrix could not distinguish the actual preferred frequencies of neurons with high $v_{pref,\ i}$. In the next section this was confirmed with alternating the deviation variables of upstream population κ (*Fig. 5* (a) & (b)). In addition, except the high frequency region, the deviation of $v_{inferred}$ increased when v_{test} was higher. Nevertheless, as the dots lad close to the line, one can still conclude that the connection matrix was trained well.

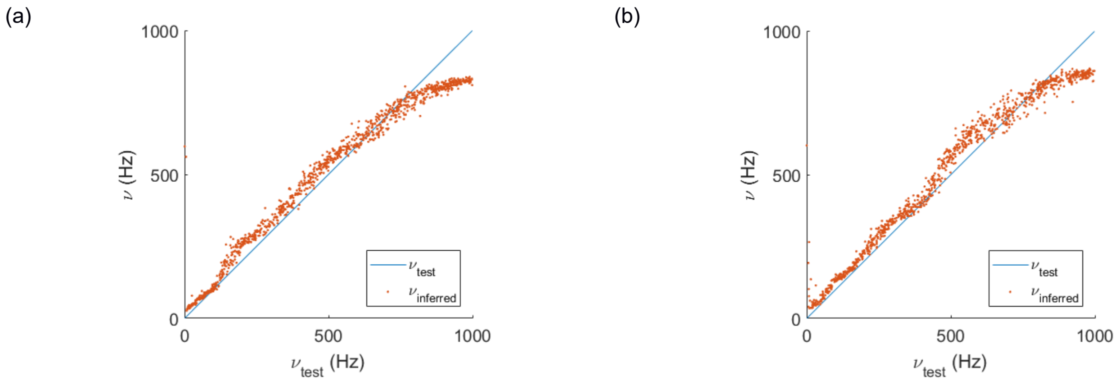


Fig. 4 Inferred stimulating frequencies resembled the actual values with basic parameter. From two different experiments (a) & (b), $v_{inferred}$ closely resembled the actual v_{test} with the exception of v > 800 *Hz*. The deviation of $v_{inferred}$ for higher v_{test} was also notably higher.

Alternating deviations of the turning curves of both populations changed the efficiency of the connection matrix

The intrinsic property of a neuron limits significantly on the formation of network constructed. To illustrate this, the deviations of the turning curves of both *M* and *N* were modified by changing values of κ and σ and the testing results were observed (*Fig. 5)*. With a drop in κ (*Fig. 5* (*a*)), the pattern appeared to be more condensed around the actual values, with occasional stair-like pattern. This is because a narrower turning curve of *N*, especially for low $v_{pref,\ i}$, the information encoded by them would be highly filtered by $v_{pref,\ i}$, and therefore $R_{test,\ j}$ would be highly dependent on the neurons with the closest $v_{pref,\ i}$ to the v_{test}. As a result, the inferred values would reflect largely the filtering effect by $v_{pref,\ i}$, as seen in the horizontal phase of the stair pattern. On the other hand, with higher κ (*Fig. 5* (b)), the pattern was much more dispersed, especially in the low frequency region of v_{test} < 300 *Hz*. Moreover, the inferred value in this region was higher than the actual one. This is because the increase in κ largely widened the turning curve of the neurons in *N* with high $v_{pref,\ i}$, and therefore small yet significant $R_{test,\ i}$ of these neurons could be triggered and misled *M*.

For a change in σ, the effect on the connection matrix was more subtle. The main effect was the hump observed with an increase in σ (*Fig. 5* (d)). The slope of this hump decreased with the increase in σ. This may reflect the fact that during the training session, with the increase in the width of the turning curve of *M*, neurons with a high $v_{pref,\ j}$ responded much stronger and thus connected more strongly with neurons with a mid-high $v_{pref,\ i}$ in *N*. Consequently, when a mid-high v_{test} was used, the connection matrix would also channel the energy to the neurons with a high $v_{pref,\ j}$, resulting in an increased inferred value of a mid-high v_{test}.

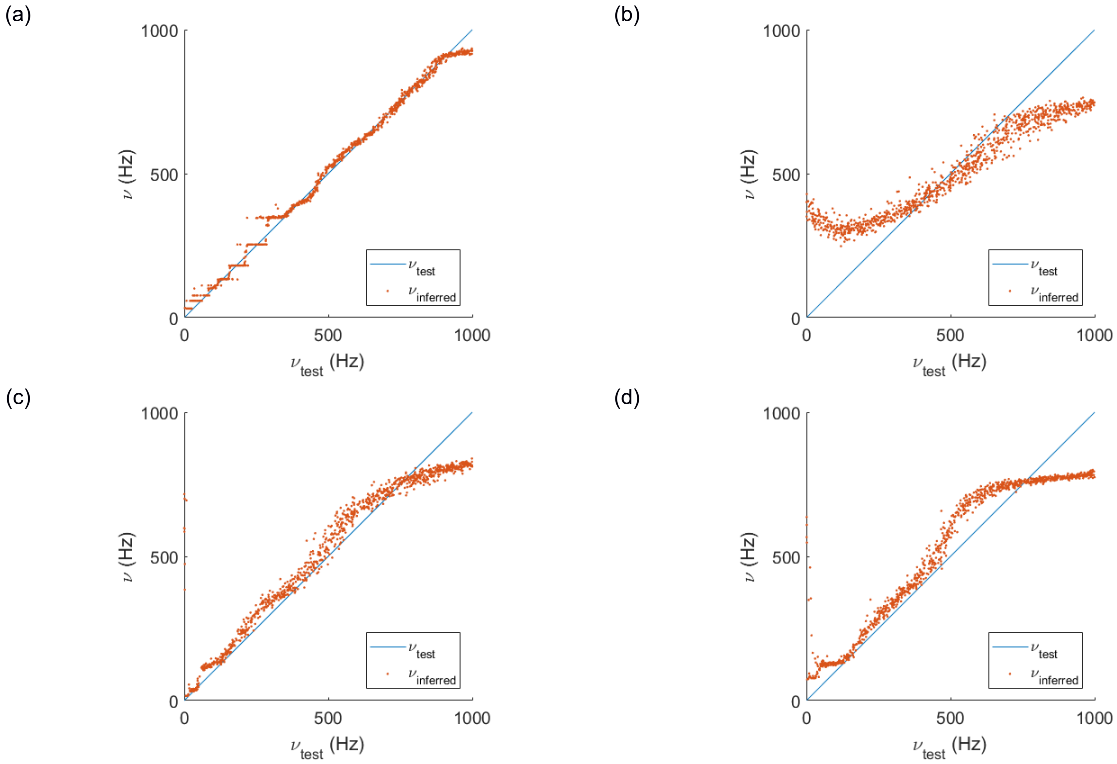


Fig. 5 Representative plots of inferred stimulating frequency against the actual one with either different κ or σ. With lower κ, the deviation of $v_{inferred}$ decreased as seen with a more condense pattern in (a) than (b), with (a) κ = 0.05 (b) κ = 0.5. With higher σ, higher deviation of $v_{inferred}$ from v_{test} could be observed in also mid-high region v_{test} between 500 to 700 *Hz* ((c) & (d)), with (c) σ = 5 (d) σ = 100.

Alternating sizes of both populations changed the efficiency of the connection matrix

With a drop in *i* (*Fig. 6* (*a*)), the pattern appeared to be more dispersed, plus an obvious hump pattern. The hump-pattern was formed mainly because the turning curve of $R_{test,\,i}$ were linear, which means that the neurons with mid-high $v_{pref,\,i}$ in *N* mistakenly connected to high $v_{pref,\,j}$ in *M*, causing the increased inferred value in mid-high region. On the other hand, an increase in *i* reduced the discrepancy between $v_{inferred}$ and v_{test} and also the deviation of $v_{inferred}$. This may because less information was masked by the linear deviation in the turning curves of *N*. Therefore a larger size of upstream population, i.e. convergence, helps better preservation of the information.

For *j*, more stair-like patterns were observed with smaller size of *j* (*Fig. 6* (a)). This is mainly because of the filtering effect of *N*, as the small population could not cover the whole range of v_{test}. No observable change appeared in the case of higher downstream population than the basic parameter (*Fig. 4*), reflecting the efficiency of the feed-forward connection was also high.

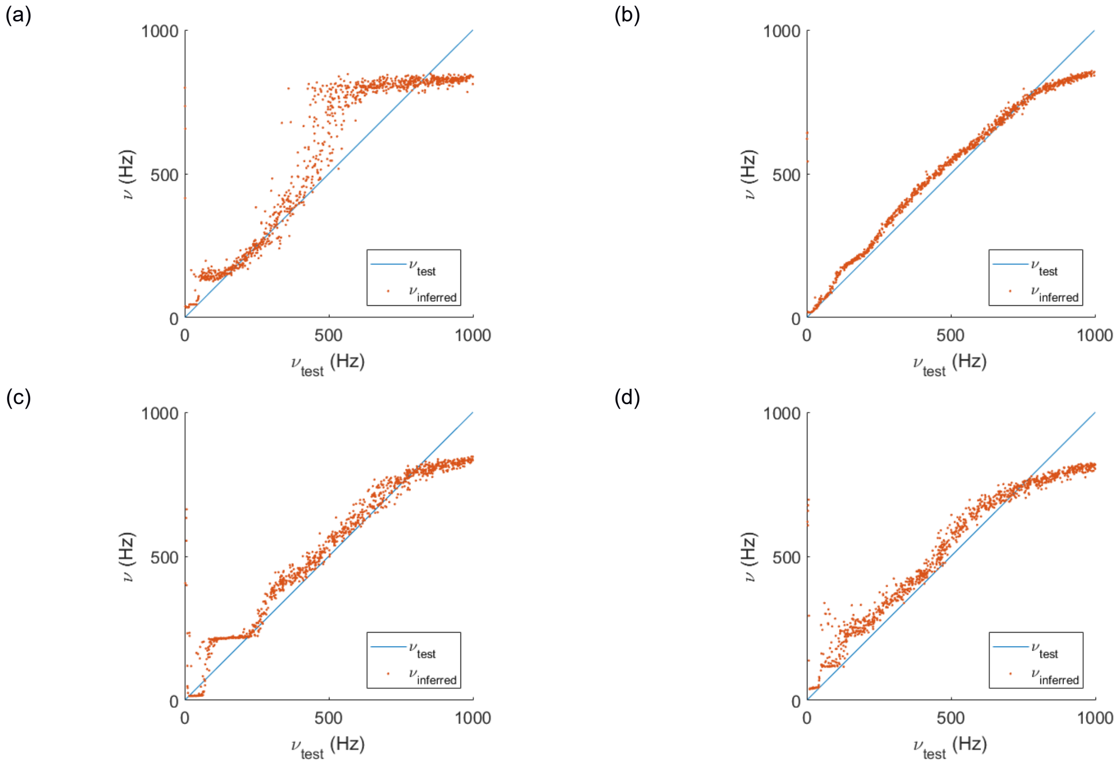


Fig. 6 Representative plots of inferred stimulating frequency against the actual one with either different i or j. With lower i, the deviation of $v_{inferred}$ increased as seen with a more disperse pattern in (a) than (b), with (a) i = 20 (b) κ = 200. In addition, a hump was also formed in lower i (a). With lower j, more stair-like pattern could be observed ((c) & (d)), with (c) σ = 20 (d) σ = 200.