1 Synaptic Weights from Receptive Fields and Stimulus Reconstruction

We consider the optimal assignments of weights to the coding of an input. We use the ”Normative Approach” to predict the input from the firing properties of the cognizant sensory neurons. This does not mean that we can perfectly reconstruct or predict the form of the input from underlying neuronal response; perception is only as good as the receptors.

We take the inputs as $I_i(t)$. This could correspond to the value of the output of a photoreceptor or, in a synthetic world, to the pixel of a camera at a time t. We take the output of a neuron as $V_i(t)$, where V is a positive number. Then

$$V_i(t) = \sum_k W_{ik} I_k(t) \tag{1.1}$$

where W_{ik} connects input I_k to the i-th neuron. This just corresponds to a linear, one level network, more properly a Perceptron. We are going to keep the clunky form of double indices and sums to maintain clarity, much as one could write $V_i(t)$ as the inner product $V_i(t) = \mathbf{W}_i \mathbf{I}(t)$.

How do we choose the W_{ik}s? We can choose them so that the $V_i(t)$s are the best predictor of the input. Then we look at the reconstruction algorithm, given by

$$I_k(t) = \sum_i W_{ik} V_i(t) \tag{1.2}$$

Let’s form a quadratic error function and minimize the argument that minimizes the error. we are going to do this in two ways. First, to get the optimal output and see if we recover our original input scheme. This will act as a form of self consistency. Second, to get the optimal form of the connection strengths, i.e., the get a rule for the online learning of new inputs.

First, let’s find the output that minimizes the error at each time point, and for each reconstructed input, where

$$\begin{aligned}
\mathrm{Error(i,k;t)} &= \| \, I_k(t) \; - \; W_{ik}(t-1)V_i(t) \, \|^2 \qquad (1.3)\\
&= \left[\| \, I_k(t) \, \|^2 \; - 2I_k(t)W_{ik}(t-1)V_i(t) + \; \| \, W_{ik}(t-1)V_i(t) \, \|^2 \right] \\
&= \left[\| \, I_k(t) \, \|^2 \; - 2I_k(t)W_{ik}(t-1)V_i(t) + \; \| \, W_{ik}(t-1) \, \|^2 \;\; V_i^2(t) \right]
\end{aligned}$$

where the time variable for the W_{ik} refers to the extent of updating.

Let’s find the argument V_i that minimizes the error, i.e., $arg\,min$. Then

$$V_i(t) \;\; = \;\; arg\,min_{V_i} \left[\sum_k \mathrm{Error(i,k;t)} \right] \tag{1.4}$$

$$
\begin{aligned}
&= arg\,min_{V_i} \sum_k \left[-2 I_k(t) W_{ik}(t-1) V_i(t) + \parallel W_{ik}(t-1) \parallel^2 V_i^2(t) \right] \\
&= arg\,min_{V_i} \sum_k \left\| \frac{\sum_k W_{ik}(t-1) I_k(t)}{\parallel \sum_k W_{ik}(t-1) \parallel^2} - V_i(t) \right\|^2 \left\| \sum_k W_{ik}(t-1) \right\|^2 \\
&= \frac{\sum_k W_{ik}(t-1) I_k(t)}{\parallel \sum_k W_{ik}(t-1) \parallel^2}.
\end{aligned}
$$

This is the same result we had previously to within a normalization. The output of each cell is the inner product of the input with the weights. So we are self-consistent.

Let's find a rule for the argument $W_{ik}(t)$ that minimizes the error, i.e., $arg\,min_{W_{i,k}}$, over all time. Then

$$
\begin{aligned}
W_{ik}(T) &= arg\,min_{W_{ik}} \left[\sum_{t=1}^{T} \mathrm{Error(i, k; t)} \right] \qquad (1.5) \\
&= arg\,min_{W_{ik}} \sum_{t=1}^{T} \left[-2 I_k(t) W_{ik}(t-1) V_i(t) + \parallel W_{ik}(t-1) \parallel^2 V_i^2(t) \right] \\
&= arg\,min_{W_{ik}} \sum_{t=1}^{T} \left\| \frac{I_k(t) V_i(t)}{V_i(t)} - W_{ik}(t-1) \right\|^2 V_i^2(t) \\
&= \frac{\sum_{t=1}^{T} I_k(t) V_i(t)}{\sum_{t=1}^{T} V_i^2(t)}.
\end{aligned}
$$

so we see that the weights are the cross-correlation of the input with the output. To get an incremental rule, we note

$$
\begin{aligned}
W_{ik}(T-1) &= \frac{\sum_{t=1}^{T-1} I_k(t) V_i(t)}{\sum_{t=1}^{T-1} V_i^2(t)} \qquad (1.6) \\
&= \frac{\sum_{t=1}^{T} I_k(t) V_i(t) - I_k(T) V_i(T)}{\sum_{t=1}^{T} V_i^2(t) - V_i^2(T)}.
\end{aligned}
$$

so

$$
\sum_{t=1}^{T} I_k(t) V_i(t) = W_{ik}(T-1) \sum_{t=1}^{T} V_i^2(t) - W_{ik}(T-1) V_i^2(T) + I_k(T) V_i(T) \qquad (1.7)
$$

Combining terms by substituting Equation 1.7 into Equation 1.5, we get

$$
\begin{aligned}
\Delta W_{ik}(T) &\equiv W_{ik}(T) - W_{ik}(T-1) \qquad (1.8) \\
&= \frac{V_i(T) \left[I_k(T) - W_{ik}(T-1) V_i(T) \right]}{\sum_{t=1}^{T} V_i^2(t)},.
\end{aligned}
$$

which is our learning rule.

The change in weight has a contribution that appears like a Hebb rule, except that the change decrements sustained plasticity. The rule depends only of pre- and post-synaptic activity and the previous value of the weight; all of this is biologically plausible. The normalization by the square of the output is worry some; presumably one must add a feature that cuts this term off after some long time so that the denominator is simply a term that depends of average (square) activity.

The learning rule is the starting point for work by Oja (1982) showing that such a rule leads to weights that approximate the first principle component of the input. Let's plug in for $V_i(T)$,

$$\begin{aligned} \Delta W_{ik}(T) &= \frac{\frac{\sum_m W_{im}(T-1)C_{mk}}{\|\sum_k W_{ik}(t-1)\|^2} - \frac{\sum_{m,n} W_{im}(T-1)C_{mn}W_{in}(T-1)}{\|\sum_k W_{ik}(t-1)\|^4} W_{ik}(T-1)}{\frac{\sum_{m,n}\sum_{t=1}^{T} W_{im}(T-1)C_{mn}W_{in}(T-1)}{\|\sum_k W_{ik}(t-1)\|^4}} \qquad (1.9) \\ &= \frac{\| \sum_k W_{ik}(T-1) \|^2 \sum_m W_{im}(T-1)C_{mk} - \sum_{m,n} W_{im}(T-1)C_{mn}W_{in}(T-1)W_{ik}(T-1)}{\sum_{m,n}\sum_{t=1}^{T} W_{im}(T-1)C_{mn}W_{in}(T-1)} \end{aligned}$$

where

$$C_{mn} = \frac{1}{T}\sum_{t=1}^{T} I_m(t) I_n(t). \qquad (1.10)$$

is the correlation matrix of the inputs and may be assumed to achieve a set of roughly constant values. Switching back to vector notation,

$$\frac{d\mathbf{W}(t)}{dt} = \frac{\| \mathbf{W}(t) \|^2 \, \mathbf{CW}(t) - \left[\mathbf{W}^T(t)\mathbf{CW}(t)\right] \mathbf{W}(t)}{\mathbf{W}^T(t)\mathbf{CW}(t)} \qquad (1.11)$$

we see that the expression is in the form of the Ojas (1982) rule for which the $\mathbf{W}$ will select the dominant eignevector of the correlation matrix or equivalently the first principle component of the correlation matrix. In steady state, $d\mathbf{W}(t)/dt = 0$, which leads to the eigenvalue equation

$$\mathbf{CW} = \frac{\mathbf{W}^T\mathbf{CW}}{\| \mathbf{W} \|^2}\,\mathbf{W} \qquad (1.12)$$

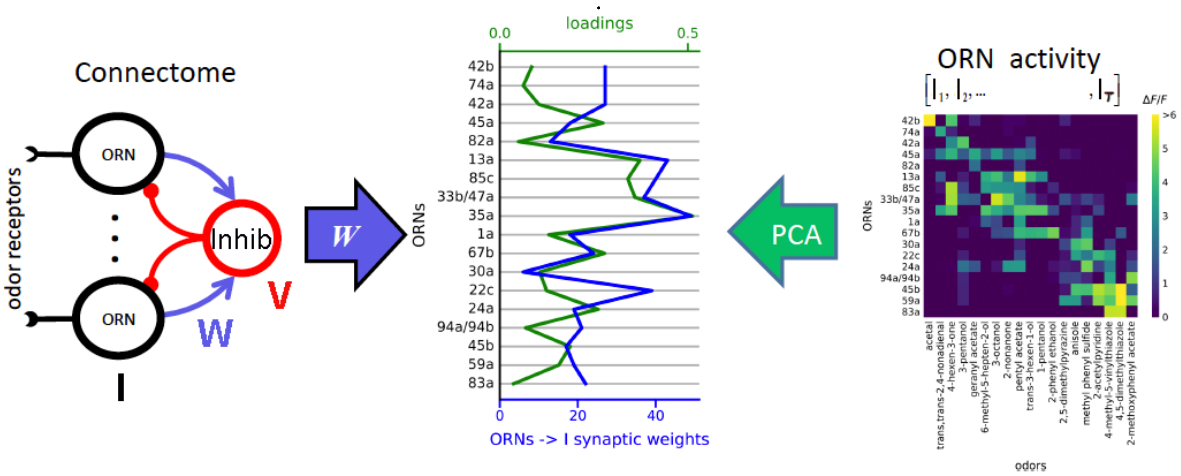
where $\mathbf{W} = \mathbf{W}(t \to \infty)$. The weights vector $\mathbf{W}$ will be dominated by the leading eigenvector of the correlation matrix of the inputs, $\mathbf{C}$. The associated eigenvalue is just

$$\frac{\mathbf{W}^T\mathbf{CW}}{\| \mathbf{W} \|^2},$$

as can be readily checked using the same approach that we used to show that a linear recurrent system can only store one memory, i.e., the dominant eigenvector,

Rather impressively, Chklovskii recently showed that this approach applied to an input space of odorants yields a matrix that matches the response of selected olfactory receptor neurons cells.

Figure 1: Observed versus calculated synaptic weights for responses in the fly From Chklovskii



Berck et al, 2016

Chapochnikov, Pehlevan, Chklovskii et al, unpublished

Si et al, 2019