Askesis: Granular Cells

Eric Purdy

July 13, 2014

1 Non-maximal suppression

Non-maximal suppression is a special case of a phenomenon called "explaining away". When there are two possible causes of a particular event, knowing that one of the causes did take place lowers the probability that the other possible cause took place. The classic example is an alarm that goes off in response to a robbery, but which can also be set off by an earthquake. Both robbery and earthquake are very rare events, so their initial probability is low. If all we know is that the alarm went off, both possible causes become more likely. If we then learn that an earthquake did indeed take place, then the probability of a robbery goes back down to something like its initial probability - the earthquake has "explained away" the evidence provided by the alarm.

Consider now the example of signal coming from muscle spindles, which fire fastest when their muscle is at a particular length. However, they also fire more quickly than normal when the muscle is close to the correct length. Consider two neurons, A, which fires most quickly when a particular muscle is 12 inches long, and B, which fires most quickly when the muscle is 13 inches long. If we then learn that both A and B are firing quickly, but that B is firing more quickly, then our interpretation should be that the muscle is thirteen inches long. The firing of A is explained away - it comes from the muscle being close to the correct length.

We can apply this logic whenever we have two similar events that are unlikely to hold simultaneously.

In order for the non-maximal suppression performed by the Golgi cell to make sense, we need to know that the granular cells suppressed by a particular Golgi cell are similar to one another. Some amount of similarity is given just by the somatotopic mapping encoded by the mossy fibers - mossy fibers that are close to one another respond to stimuli from regions of the body that are close to one another.

We thus postulate that the granular cells under a particular Golgi cell function somewhat like a flock of birds: at each time step, each granular cell updates its parameters to be more correlated with the Golgi cell, which can be thought of as the center of the flock. The overall effect is thus to create an ensemble of granular cells to which non-maximal suppression can be meaningfully applied.

2 Maximizing the Covariance

Recall the logistic function

$$\sigma(x) = \frac{1}{1+e^{-x}}.$$

Recall that its derivative is

$$\frac{d}{dx}\sigma(x) = \frac{e^{-x}}{(1+e^{-x})^2} = \sigma(x)(1-\sigma(x)).$$

Consider the situation where we have a number of granular cells which all excite the same Golgi cell, and which receive inhibitory signals from that Golgi cell.

Let x_i be a binary variable (either 0 or 1) that encodes whether the i-th mossy fiber fired at a particular time step. Let x be the vector of all the mossy fiber activations, so that x_i is 1 exactly when the i-th mossy fiber fires, and 0 otherwise. Let $p(x)$ be the probability of observing the value x. We will assume that we receive a sample from this distribution at every time step. In this section, we will assume that each sample is independent of the others.

Let the i-th granular cell's output be defined by

$$G_i(x) = \sigma\left(\sum_j w_{ji}x_j - \theta_i\right).$$

Let the Golgi cell's output be defined by

$$Z(x) = \sigma\left(\sum_i G_i(x) - \varphi\right).$$

Note that we are assuming that the Golgi cell weights all of its inputs evenly.

As discussed above, we wish to maximize the covariance between G_i and Z.

$$\begin{aligned}
Cov(G_i, Z) &= \mathbb{E}_x[G_i(x)Z(x)] - \mathbb{E}_x[G_i(x)]\mathbb{E}_y[Z(y)] \\
&= \sum_x p(x)G_i(x)Z(x) - \sum_x p(x)G_i(x)\sum_y p(y)Z(y) \\
\frac{\partial Cov(G_i, Z)}{\partial w_{ji}} &= \sum_x p(x)\left[\frac{\partial G_i(x)}{\partial w_{ji}}Z(x) + G_i(x)\frac{\partial Z(x)}{\partial w_{ji}}\right] \\
&\quad - \sum_x p(x)G_i(x)\sum_y p(y)\frac{\partial Z(y)}{\partial w_{ji}} - \sum_x p(x)\frac{\partial G_i(x)}{\partial w_{ji}}\sum_y p(y)Z(y)
\end{aligned}$$

Letting $\overline{Z} = \sum_y p(y)Z(y)$ and $\overline{G_i} = \sum_x p(x)G_i(x)$, we have

$$
\begin{aligned}
&= \sum_x p(x) \left[\frac{\partial G_i(x)}{\partial w_{ji}} Z(x) + G_i(x) \frac{\partial Z(x)}{\partial w_{ji}} \right] \\
&\qquad - \overline{G_i} \sum_x p(x) \frac{\partial Z(x)}{\partial w_{ji}} - \overline{Z} \sum_x p(x) \frac{\partial G_i(x)}{\partial w_{ji}} \\
&= \sum_x p(x) \left(Z(x) - \overline{Z} \right) \frac{\partial G_i(x)}{\partial w_{ji}} + \sum_x p(x) \left(G_i(x) - \overline{G_i} \right) \frac{\partial Z(x)}{\partial w_{ji}}
\end{aligned}
$$

Now, by the chain rule

$$
\begin{aligned}
\frac{\partial G_i(x)}{\partial w_{ji}} &= \sigma' \left(\sum_k w_{ki} x_k - \theta_i \right) \cdot \frac{\partial}{\partial w_{ji}} \left(\sum_k w_{ki} x_k - \theta_i \right) \\
&= G_i(x)(1 - G_i(x)) x_j
\end{aligned}
$$

and similarly

$$
\begin{aligned}
\frac{\partial Z}{\partial w_{ji}}(x) &= Z(x)(1 - Z(x)) \frac{\partial}{\partial w_{ji}} \left(\sum_k G_k(x) - \phi \right) \\
&= Z(x)(1 - Z(x)) \frac{\partial G_i(x)}{\partial w_{ji}} \\
&= Z(x)(1 - Z(x)) G_i(x)(1 - G_i(x)) x_j
\end{aligned}
$$

Substituting this into the above, we arrive at

$$
\begin{aligned}
\frac{\partial Cov(G_i, Z)}{\partial w_{ji}} &= \sum_x p(x)(Z(x) - \overline{Z}) G_i(x)(1 - G_i(x)) x_j \\
&\qquad + \sum_x p(x)(G_i(x) - \overline{G_i}) Z(x)(1 - Z(x)) G_i(x)(1 - G_i(x)) x_j \\
&= \sum_x p(x) x_j G_i(x)(1 - G_i(x)) \left[Z(x) - \overline{Z} + (G_i(x) - \overline{G_i}) Z(x)(1 - Z(x)) \right] \\
&= \sum_x p(x) x_j G_i(x)(1 - G_i(x)) Z(x)(1 - Z(x)) \left[G_i(x) - \overline{G_i} + \frac{Z(x) - \overline{Z}}{Z(x)(1 - Z(x))} \right]
\end{aligned}
$$

We can use stochastic gradient descent to maximize the covariance, yielding the update rule

$$
\Delta w_{ji} = x_j G_i(x)(1 - G_i(x)) Z(x)(1 - Z(x)) \left[G_i(x) - \overline{G_i} + F(Z) \right],
$$

where $F(Z) = \frac{Z - \overline{Z}}{Z(1 - Z)}$.

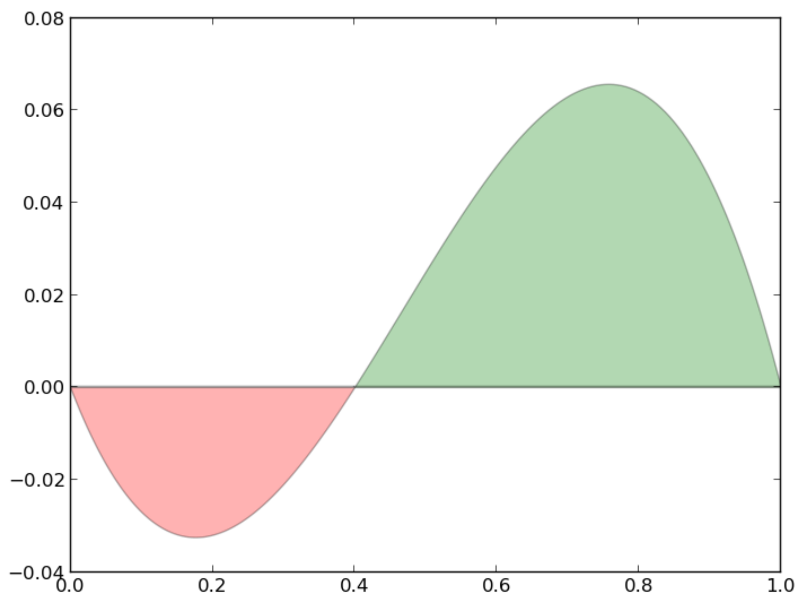

Figure 1: Δw_{ji} as a function of G_i, with $(\overline{G_i} - A) = 0.4$.

For a fixed Z, we then have (letting $A = F(Z)$)

$$\Delta w_{ji} \propto x_j G_i(x)(1 - G_i(x))(G_i(x) - (\overline{G_i} - A)),$$

which yields the BCM rule, with zeros at $G_i = 0, G_i = 1$, and $G_i = \overline{G_i} - A$. Note that the middle threshold, where Δw_{ji} switches from negative to positive, varies as a function of Z. Changing Z will thus switch the sign of Δw_{ji} for values of G close to the middle threshold.

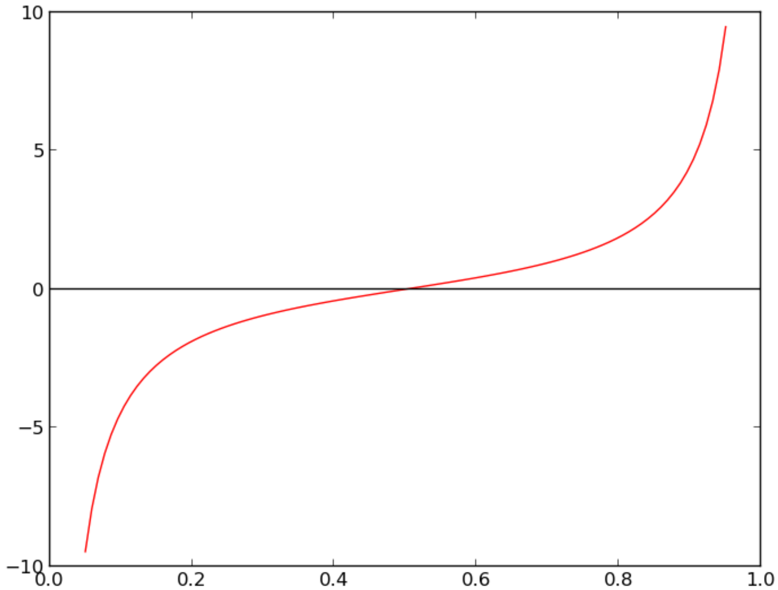

Figure 2: $F(Z) = \frac{Z - \overline{Z}}{Z(1-Z)}$, with $\overline{Z} = 0.5$