

Assignment 8

Theoretical Neuroscience

Peter Dayan, Quentin Huys

28 November 2004

1. BCM and projection pursuit

We have seen in class that straightforward application of the Hebbian learning rule will lead to a weight vector parallel to the leading eigenvector of the correlation matrix or of the covariance matrix if the training patterns have zero mean.

The BCM learning rule

$$\tau \frac{d\mathbf{w}}{dt} = \mathbf{u}v(v - \theta) \quad (1)$$

$$\tau_\theta \frac{d\theta}{dt} = -\theta + v^2 \quad (2)$$

will lead to somewhat more interesting solutions.

A Produce N d -dimensional Gaussian samples \mathbf{u} (e.g. $N = 5000$, $d = 50$). Add a dimension to this that is non-Gaussian (do it once for a bimodal distribution and once for heavy tails). Look at the projection of these samples onto 20 random vectors \mathbf{p}_i on the unit hypersphere. What do the distributions of these projections look like?

B Projection Pursuit is based on the observation that most projections of most high-dimensional data produce approximated Gaussian distributions. The most 'interesting' (exceptional) projections are thus those that look least Gaussian. Projection pursuit defines a 'projection index' – some cost function that defines which projected distributions are of interest. A sensible, general, projection index might for example be the KL divergence between the projected distribution and a Gaussian with the same mean and variance. For the data set and the vectors \mathbf{p}_i from above which is the vector that produces the least Gaussian projection in a KL sense? Is it what you expected?

C The BCM learning rule can be related to projection pursuit in that it finds some 'interesting' aspect of the data. To see why this is the case, consider the cost function

$$C_{\mathbf{w}}(\mathbf{u}) = -\alpha \int_0^{\mathbf{w} \cdot \mathbf{u}} s^2 - \frac{1}{2} s \Theta_{\mathbf{w}} ds \quad (3)$$

$$\Theta_{\mathbf{w}} = \langle (\mathbf{u} \cdot \mathbf{w})^2 \rangle_{\mathbf{u}}. \quad (4)$$

What kind of distributions of v will this cost function penalise / allow? Hint: Imagine a distribution with two modes. Where would you put the modes? Where would you put the mode of a heavy-tailed distribution?

D OPTIONAL Now let us do gradient descent on the expectation of this loss:

$$\frac{d\mathbf{w}}{dt} = -\frac{dR}{d\mathbf{w}} \quad (5)$$

$$\text{where } R = \langle C_{\mathbf{w}}(\mathbf{u}) \rangle \quad (6)$$

Show that this gives a learning rule akin to equation 1, with equation 2 replaced by 4.

E Apply this learning rule to the data above, and show that it picks out the statistically 'interesting' direction... or does it? Compare different bimodal directions. Does it pick the one producing a projection more distant from a Gaussian in terms of KL divergence?

F OPTIONAL Now assume you have multiple postsynaptic neurones. To prevent them from doing the same, let them inhibit each other:

$$\bar{v}_i = v_i - \eta \sum_{k \neq i} v_k.$$

Derive the learning rule for this case and apply it to the above problem, but with two non-Gaussian dimensions.

2. STDP and predictive Hebbian learning

Consider a collection of one-dimensional hippocampal place cells with mean firing rates $r_y(x) = \exp(-(x - y)^2/2\sigma^2)$, where the spatial position of the rat is x , and (to abuse notation) cell y has preferred location y . Assume that decoding can be based on the mean firing rates, and work in the infinite limit of dense coverage.

A Work out $\int dy r_y(x)y$ and $\int dy r_y(x)$, and so show that

$$x = \hat{x}(\mathbf{r}) \equiv \frac{\int dy r_y(x)y}{\int dy r_y(x)}$$

B If cell y' is connected to cell y with a small weight $W(y', y)$, so that the net activity

$$r_y^o(x) = r_y(x) + \int W(y, y') r_{y'}(x) dy'$$

show that, to first order in W , when the rat is at x ,

$$\hat{x}(\mathbf{r}^o) = \frac{\int dy r_y^o(x)y}{\int dy r_y^o(x)} \approx x + \frac{1}{\int dy r_y(x)} \int (y - x) W(y, y') r_{y'}(x) dy dy' \quad (7)$$

C If the weights are set in an idealised predictive manner, based on the rat running from left to right, so that

$$W(y, y') = \begin{cases} 0 & \text{if } y' < y \\ \exp(-(y' - y)/\lambda) & \text{if } y' \geq y \end{cases}$$

D OPTIONAL Calculate the displacement to the right that is implied by the approximate decoding in equation 7

3. Essay OPTIONAL

One can argue that early sensory systems should try to encode information as efficiently as possible. What are the biochemical / biophysical mechanisms that make photoreceptor good 'channels'?