

Lecture 12.3

- ▶ This lecture describes how linear filters can be learned from images by unsupervised algorithms or estimated from neural data by regression. We describe how these receptive field models can be used for binocular stereo and for motion estimation.
- ▶ Then we introduce probabilities and decision theory. We motivate this by discussing how cues can be combined to detect edges in images.
- ▶ This lecture includes exercises involving interactive demos: (12.3.1) Oja's Rule and Principal Component Analysis, (12.3.2) Natural Image Statistics, and (12.3.3) Statistical Edge Detection.

Unsupervised learning of the receptive fields.

- ▶ We now introduce unsupervised neural network algorithms for learning receptive fields. This section is based on computational studies performed in the 1980's (Linsker, 1986a,b; Yuille et al., 1989), see (Zhaoping, 2014) for other references. These studies are based on modifications of the Hebb learning rule, which has some experimental support. Exercise demo (12.3.1) illustrates principal component analysis and Oja's rule (Oja, 1982).
- ▶ The basic findings are that center-surround, orientation selective, quadrature pairs, and disparity sensitive cells (precursors to cells that can estimate depth from binocular stereo) could all be obtained by variants of the same learning rule. Analysis of these findings suggest that this is partly due to the shift invariance of images.

Unsupervised learning by Hebb's rule (I)

- ▶ We first describe a simple unsupervised learning model for a single cell (Oja, 1982). The output $S(t)$ of the cell is a function of time t and is a weighted sum of the inputs $I_i(t)$, where the weights $w_i(t)$ are functions of time and are updated by Oja's rule (Oja, 1982):

$$S(t) = \sum_j w_j(t) I_j(t),$$

$$\frac{dw_i(t)}{dt} = S(t) \{I_i(t) - S(t)w_i(t)\}. \quad (7)$$

- ▶ The first term (Hebbs) increases the strength of a weight w_i if its input $I_i(t)$ is positively correlated with the output $S(t)$ (i.e., $\langle S(t)I_i(t) \rangle > 0$), while the second term decreases the value of all weights by an amount proportional to their strength.
- ▶ This can be expressed as a single update equation:

$$\frac{dw_i(t)}{dt} = \sum_j w_j I_i(t) I_j(t) - \sum_{jk} w_i w_j w_k I_j(t) I_k(t). \quad (8)$$

Unsupervised learning by Hebb's rule: Analysis (I)

- ▶ Next we assume that the weights w_i change at a slower rate than the input images. This enables us to replace the terms $I_i(t)I_j(t)$ with their expectation $K_{ij} = \langle I_i(t)I_j(t) \rangle$, which is the correlation function of the input. This gives:

$$\frac{dw_i(t)}{dt} = \sum_j w_j K_{ij} - \sum_{jk} w_i w_j w_k K_{jk}. \quad (9)$$

- ▶ The fixed points of this equation, the values of w such that $\frac{dw_i(t)}{dt} = 0$, can be shown to be eigenvectors of the correlation function K_{ij} . A slight modification gives an update rule (Yuille et al., 1989) that converges to the global minimum of the cost function:

$$E(\vec{w}) = -(1/2) \sum_{i,j} K_{ij} w_i w_j + (k/4) \left(\sum_i w_i^2 \right)^2$$

Unsupervised learning by Hebb's rule: Analysis (II)

- ▶ The global minimum corresponds to the biggest eigenvalue of K_{ij} . If the correlation function K_{ij} decreases with distance, then the biggest eigenvalue is at frequency 0, so the cell is not tuned to any frequency. But if the correlation function has the shape of a Mexican hat, then the biggest eigenvalue has a nonzero frequency, which implies that the cell is orientated (Yuille et al., 1989).
- ▶ The correlation function of natural images does decrease spatially, but Linsker (1986a,b) showed that correlation functions similar to the Mexican hat arise if this learning procedure is applied to a sequence of layers.
- ▶ This analysis yields receptive fields that are sinusoids, and hence have no spatial fall-off, which is unrealistic. But receptive fields of neurons are limited by the geometrical positions of the dendrites. If these constraints are included, then the algorithms converge to receptive fields that are similar to Gabor functions.

How to empirically estimate receptive field models by regression.

- ▶ We can estimate the receptive field properties of cells from electrical recordings of neurons by estimating the best model using *regression*. This makes few assumptions about the form of the receptive field.
- ▶ Recall that the receptive field properties of neurons are traditionally found by probing their response to different perceptual dimensions, such as orientations and frequency. This gives a classification of the type of the receptive field but does not specify its receptive field weights \vec{w} unless strong assumptions are made (e.g., that the receptive field is a Gabor function).

Estimating receptive field models by regression.

- ▶ The regression method makes few assumptions about the forms of the receptive field, but it does require more data. It requires a stimulus data set of $\mathcal{S} = \{(S^\mu, \vec{I}^\mu) : \mu = 1, \dots, N\}$ of inputs \vec{I}^μ and outputs S^μ (e.g., the firing rates). It requires a model, such as $g(\vec{I} : \vec{w}) = \sigma(\vec{w} \cdot \vec{I})$, where $\sigma(\cdot)$ is a sigmoid function.
- ▶ Regression requires minimizing a cost function like:

$$F(\vec{w}) = \frac{1}{|\mathcal{S}|} \sum_{\mu \in \mathcal{S}}^N E(S^\mu - g(I^\mu; \vec{w}))$$

where $E(\cdot)$ is a penalty function, e.g., $(S^\mu - g(I^\mu; T))^2$.

- ▶ This minimization can be done by standard computer packages. It outputs an estimate of the model parameters \vec{w}^* and an error measure $F(\vec{w}^*) = \frac{1}{|\mathcal{S}|} \sum_{\mu \in \mathcal{S}} E(S^\mu - g(I^\mu; \vec{w}^*))$.

Complications (I)

In practice, there are several complications. It is unrealistic to show the neuron all possible stimuli because there are so many possible image stimuli. Hence researchers have to choose a restricted set of stimuli. If neurons are linear, or a nonlinear function of a linear filter, then this should not matter because we can exploit the superposition principle and estimate the receptive field from a limited number of stimuli. But in reality, linearity is only an approximation, and in practice, the choice of stimuli can matter considerably. One concern is that the stimulus set does not contain the types of stimuli that the neuron is most sensitive to, in which case regression will output unreliable estimates. Also, if the linear assumption is only partially correct, then there is no guarantee that the receptive field learned on one set of stimuli will predict the behavior well on another set of stimuli.

Complications (II)

The complications are illustrated by recent findings (Talebi & Baker, 2012) that estimates of the receptive fields of neurons can depend heavily on the set of stimuli. The authors used three different stimulus sets: (1) white noise (WN), (2) oriented bars (B), and (3) natural images (NI). This gives three estimates for the receptive fields \vec{w}_{WN} , \vec{w}_B , \vec{w}_{NI} by using stimulus sets \mathcal{S}_{WN} , \mathcal{S}_B , \mathcal{S}_{NI} . For each data set, they compute the prediction errors F_{WN} , F_B , F_{NI} which are the errors for that data set, e.g., $F_{WN}(\vec{w}_{WN}^*) = \frac{1}{|\mathcal{S}_{WN}|} \sum_{\mu \in \mathcal{S}_{WN}} E(S^\mu - g(I^\mu; \vec{w}_{WN}^*))$. These quantities show how well the models can fit each stimulus set. They can also enable us to study how well the estimated receptive field from one stimulus set can predict the other data sets. This involves computing quantities such as $F_{WN}(\vec{w}_B^*)$, $F_{WN}(\vec{w}_{NI}^*)$, $F_B(\vec{w}_{WN}^*)$, $F_{WN}(\vec{w}_{NI}^*)$, $F_{NI}(\vec{w}_{WN}^*)$, $F_{WN}(\vec{w}_B^*)$. They show that the receptive fields estimated on the natural image stimulus set were much better at predicting the responses on the other two stimulus sets.