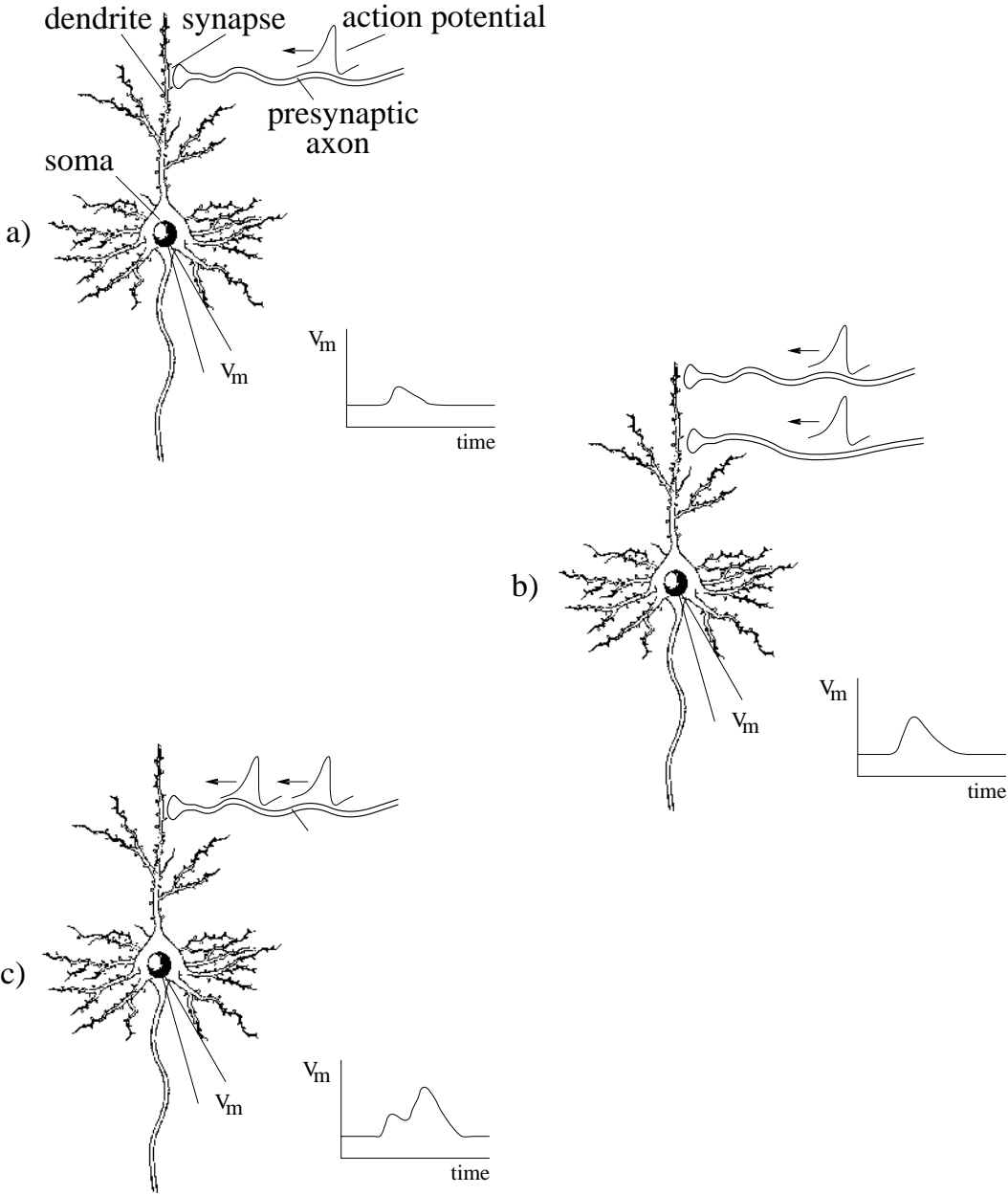

An Energy Formulation of Learning:
how neurons decide what's important

BRIAN BLAIS
PROF. LEON N COOPER

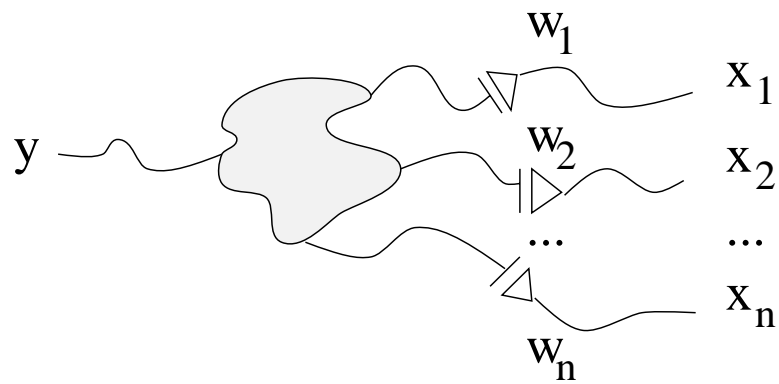
Outline

- Introduction to Neurons
- Experimental Foundation
- What are Learning Rules?
 - Energy Formulation
- Natural Scenes
- A Simplified Environment
 - One Dimensional Case
 - Two Dimensional Case
- Conclusions

Neuron

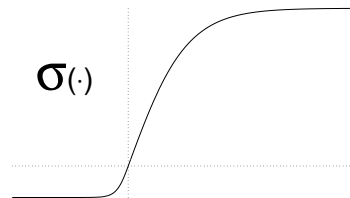


Model Neuron

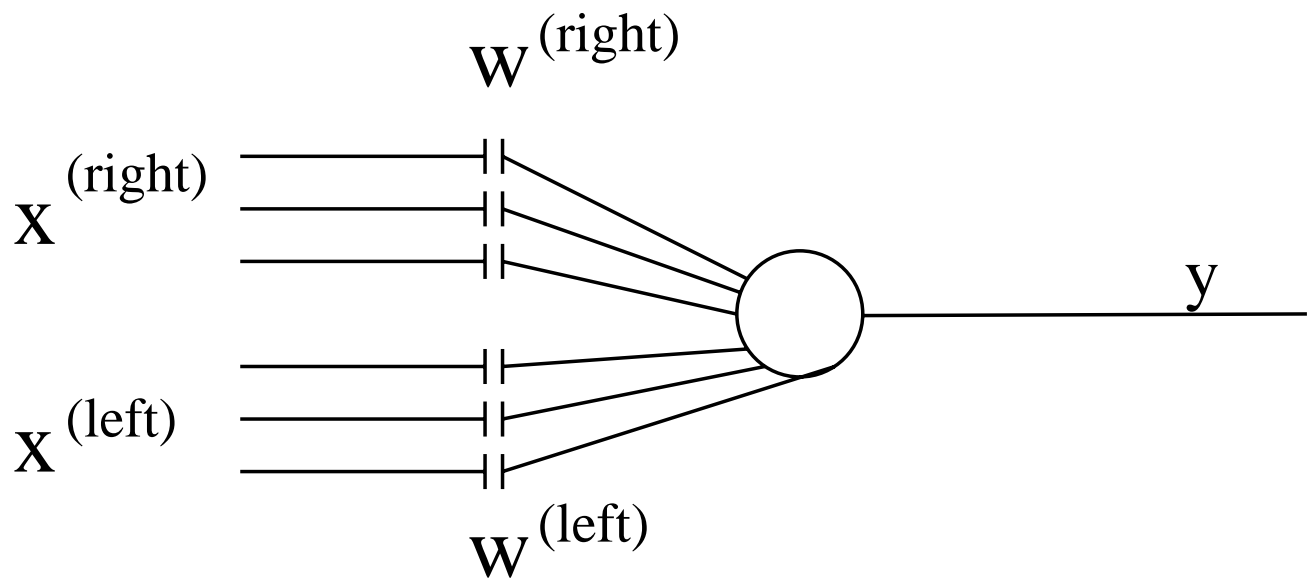
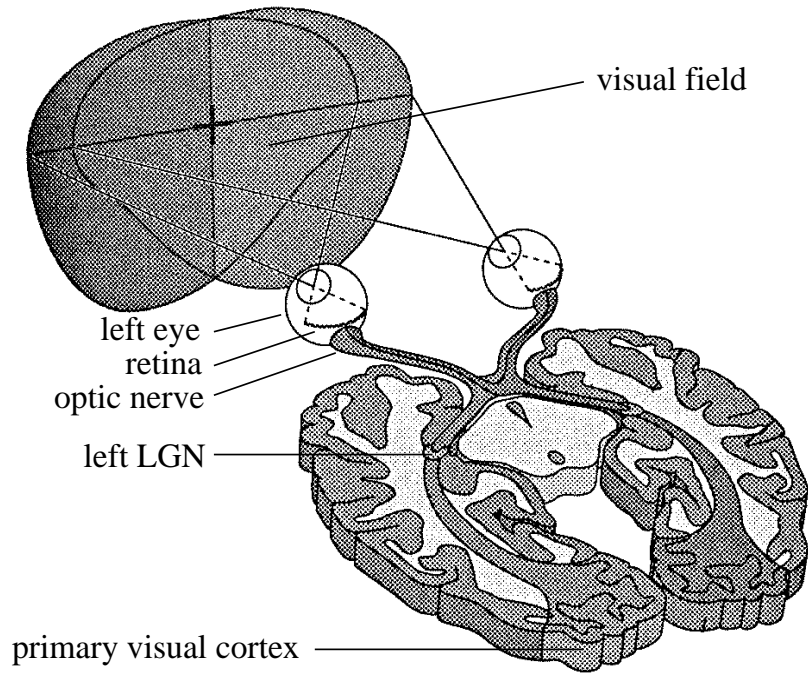


- input signals in a vector \mathbf{x}
- synaptic weight vector \mathbf{w}
- output signal scalar $y = \sigma(\mathbf{w} \cdot \mathbf{x})$,

where $\sigma(\cdot)$ looks like



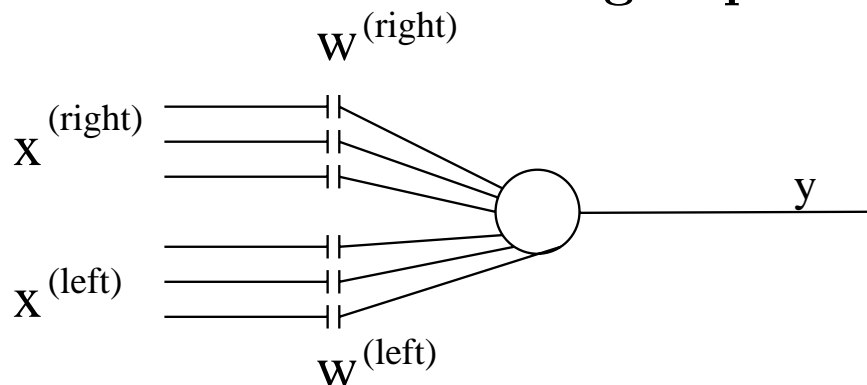
Visual System



Experimentally Modifying the Environment

- Cortical neurons are
 - *orientation selective*: respond to bars of light of a particular orientation
 - *binocular*: respond equally well to both eyes
- Monocular Deprivation (close one eye)
 - normal cells “learn” to become responsive to the open eye
 - not just a variance change!
- Stripe Rearing and Deprivation
 - cells are biased towards a particular orientation, or away from a particular orientation

What are neurons finding important?



Hebb Rule

When an axon in cell A is near enough to excite cell B and repeatedly and persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency in firing B, is increased.

$$\dot{\mathbf{w}} = y\mathbf{x}$$

- Not stable!

$$\begin{aligned} \dot{\mathbf{w}} &= y\mathbf{x} \text{ with constraint } \|\mathbf{w}\|^2 = 1 \\ \text{or } \dot{\mathbf{w}} &= y\mathbf{x} - y^2\mathbf{w} \end{aligned}$$

- What does this learning rule do?

$$\begin{aligned} \text{(fixed points) } E[\dot{\mathbf{w}}] &= 0 \\ &= E[\mathbf{x}(\mathbf{x} \cdot \mathbf{w}) - (\mathbf{x} \cdot \mathbf{w})^2\mathbf{w}] \\ &= E[(\mathbf{xx}^T)\mathbf{w} - (\mathbf{w}^T\mathbf{xx}^T\mathbf{w})\mathbf{w}] \end{aligned}$$

$$\mathbf{C} \equiv E[\mathbf{xx}^T] \Rightarrow$$

$$E[\dot{\mathbf{w}}] \equiv \mathbf{C}\mathbf{w} - (\mathbf{w}^T\mathbf{C}\mathbf{w})\mathbf{w}$$

$$\mathbf{C}\mathbf{w} = \lambda\mathbf{w}$$

$$\begin{aligned} \lambda &= \mathbf{w}^T\mathbf{C}\mathbf{w} = \mathbf{w}^T\lambda\mathbf{w} = \lambda|\mathbf{w}|^2 \\ \Rightarrow |\mathbf{w}|^2 &= 1 \end{aligned}$$

Energy Formulation

- Projection Pursuit

- Neuron searches for a *direction* in the space which maximizes some statistic of the projection, like variance
- output is the projection of the input onto the weights

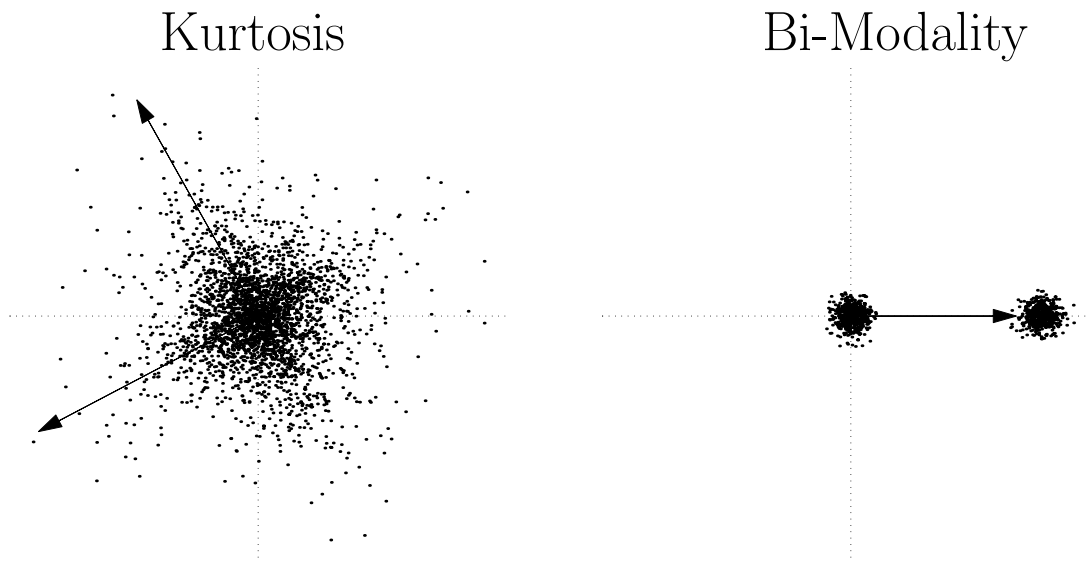
$$y = \mathbf{x} \cdot \mathbf{w}$$

- weight vector is the direction being modified to maximize energy function

- Example with Variance

$$\begin{aligned} V &= \frac{1}{2} E [y^2] \\ &= \frac{1}{2} E [(\mathbf{x} \cdot \mathbf{w})^2] \\ \frac{d\mathbf{w}}{dt} &\equiv \nabla_{\mathbf{w}} V \\ &= E [y\mathbf{x}] \end{aligned}$$

Possible Energy Functions



- Variance $V = \frac{1}{2}E[y^2]$
- Stabilized Variance $V_{\text{PCA}} = \frac{1}{2}E[y^2] \left(1 - \frac{1}{2}\mathbf{w}^2\right)$
- Bi-Modality $R_{\text{BCM}} = \frac{1}{3}E[y^3] - \frac{1}{4}E^2[y^2]$
- Kurtosis $K_1 = E[y^4] / E^2[y^2] - 3$
- Kurtosis $K_2 = E[y^4] - 3E^2[y^2]$

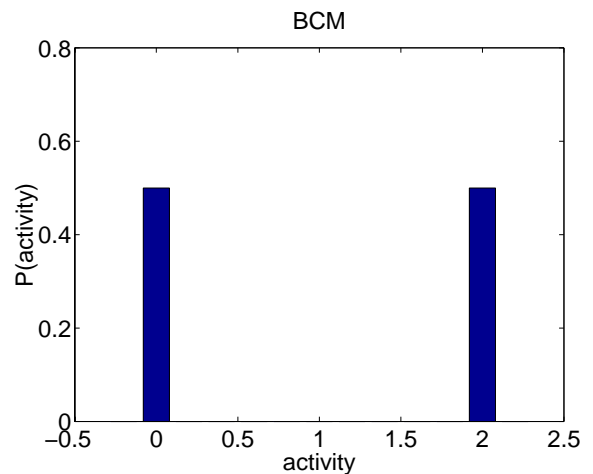
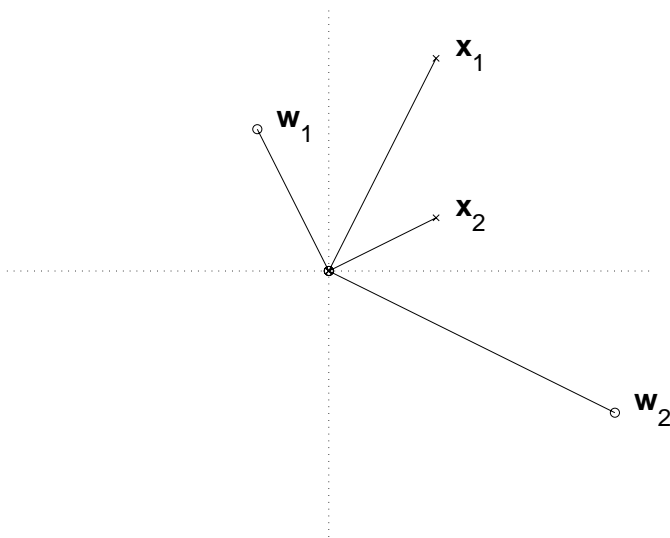
BCM Learning Rule

$$R_{\text{BCM}} = \frac{1}{3}E[(\mathbf{x} \cdot \mathbf{w})^3] - \frac{1}{4}E^2[(\mathbf{x} \cdot \mathbf{w})^2]$$

$$\frac{d\mathbf{w}}{dt} \equiv \nabla_{\mathbf{w}} R_{\text{BCM}}$$

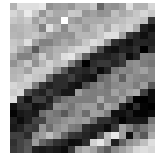
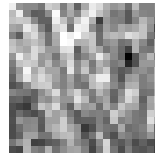
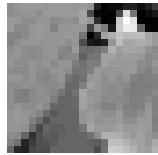
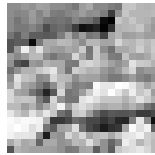
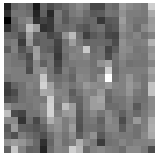
$$= E[y(y - E[(\mathbf{x} \cdot \mathbf{w})^2])\mathbf{x}] \equiv E[y(y - \theta)]\mathbf{x}$$

- two input patterns, \mathbf{x}_1 and \mathbf{x}_2
- fixed $E[\dot{\mathbf{w}}] = 0$, or either $y = 0$ or $y = \theta$
 - $(\mathbf{w} \cdot \mathbf{x}_1) = 0$ and $(\mathbf{w} \cdot \mathbf{x}_2) = 0$ (unstable)
 - $(\mathbf{w} \cdot \mathbf{x}_1) = 0$ and $(\mathbf{w} \cdot \mathbf{x}_2) = \theta \neq 0$
 - $(\mathbf{w} \cdot \mathbf{x}_1) = \theta \neq 0$ and $(\mathbf{w} \cdot \mathbf{x}_2) = 0$
 - $(\mathbf{w} \cdot \mathbf{x}_1) = \theta \neq 0$ and $(\mathbf{w} \cdot \mathbf{x}_2) = \theta \neq 0$ (unstable)

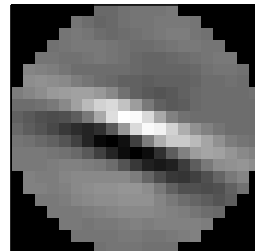
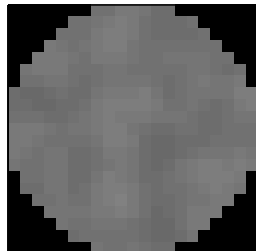


Learning Orientation Selectivity

- patches from natural scenes are presented to the neuron

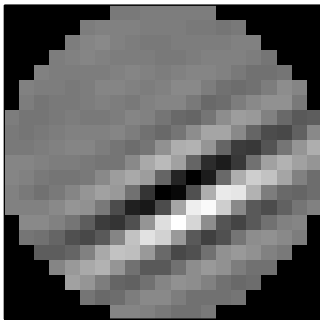


- each pixel of the receptive field is a weight
 - white denotes high value
 - black denotes low value

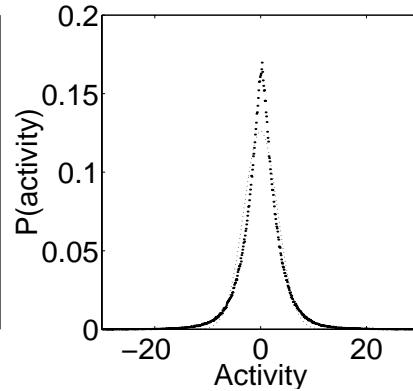


Distribution from Natural Scenes

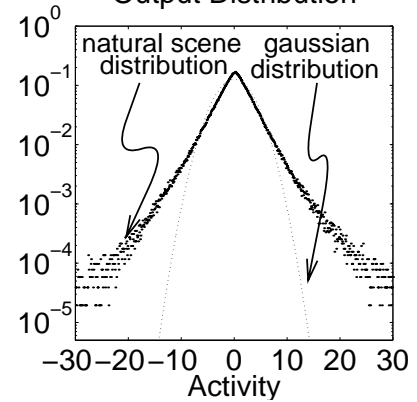
Example Receptive Field



Output Distribution



Output Distribution

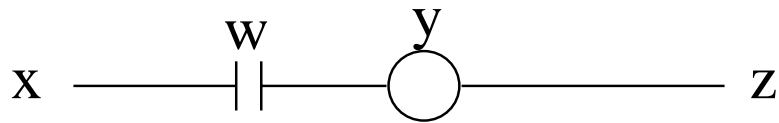


- the weight vector, and the input patches, can be thought of as vectors in a 100 dimensional space
- output distribution is
 - probability of a pattern in the environment giving a particular response
 - histogram of the projection values, when the input points, in 100D, are projected onto the weight vector
- clearly not bi-modal, or gaussian
- looks like a double exponential, or $e^{-|y|}$

Finding a Simpler Environment

- assume a distribution for the inputs
- calculate a distribution for outputs
- calculate the energy function from moments
- from gradient of the energy function, determine weight fixed points and dynamics

1D Example



$$\text{(assume)} \quad f_x(x) = \frac{1}{2\lambda} e^{-|x|/\lambda}$$

$$y = x \cdot w \quad (\text{note: no sigmoid})$$

$$z \equiv \sigma(y) \equiv \begin{cases} y & \text{if } y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$f_y(y) = \frac{1}{2\lambda|w|} e^{-|y/w|/\lambda}$$

$$f_z(z) = \begin{cases} f_y(y) & \text{if } y > 0 \\ \frac{1}{2}\delta(y) & \text{if } y = 0 \\ 0 & \text{if } y < 0 \end{cases}$$

$$\text{(assume)} \quad w = |w|$$

1D Example (continued)

- Energy Function

$$\begin{aligned} R_{\text{BCM}} &= \frac{1}{3}E[z^3] - \frac{1}{4}E^2[z^2] \\ &= \frac{1}{3} \int_{-\infty}^{\infty} z^3 f_z(z) dz - \frac{1}{4} \left(\int_{-\infty}^{\infty} z^2 f_z(z) dz \right)^2 \\ &= w^3 \lambda^3 - \frac{1}{4} w^4 \lambda^4 \end{aligned}$$

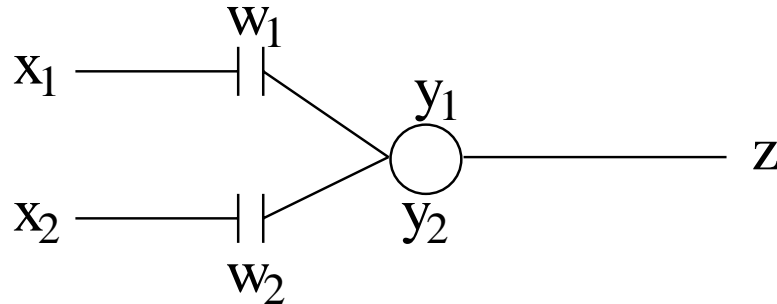
- Fixed Points

$$\begin{aligned} \frac{dR}{dw} &= 3w^2 \lambda^3 - w^3 \lambda^4 \\ &= w^2 \lambda^3 (3 - w\lambda) = 0 \\ \Rightarrow w &= 0, \frac{3}{\lambda} \end{aligned}$$

- Stability

$$\begin{aligned} \frac{d^2 R}{dw^2} &= 6w \lambda^3 - 3w^2 \lambda^4 \\ &= 3w \lambda^3 (2 - w\lambda) \\ &= \begin{cases} \frac{d^2 R}{dw^2} \Big|_{w=0} = 0 & \text{unstable} \\ \frac{d^2 R}{dw^2} \Big|_{w=3/\lambda} < 0 & \text{stable} \end{cases} \end{aligned}$$

2D Model (Laplace-Laplace)



- this is *not* the 2 eye model
- model is 2D *each* eye, or 4D
 - will reduce to slightly different 1D each eye, or 2D, problem
- “natural environment”: laplace (double exponential)
- closed eye “noise”: either uniform or gaussian

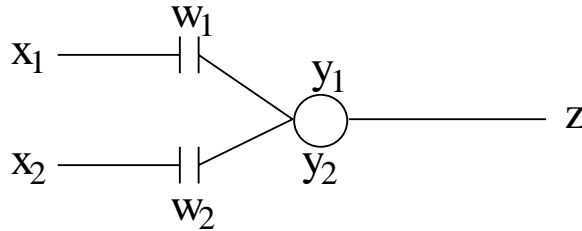
$$y = \mathbf{x} \cdot \mathbf{w} = x_1 w_1 + x_2 w_2 \equiv y_1 + y_2$$

$$z \equiv \sigma(y)$$

$$f_{x_i}(x_i) = \frac{1}{2\lambda} e^{-|x_i|/\lambda}, \quad f_{y_i}(y_i) = \frac{1}{2\lambda w_i} e^{-|y_i|/w_i \lambda}$$

$$\begin{aligned} f_y(y) &= \int_{-\infty}^{-\infty} f_{y_1}(y - y_2) f_{y_2}(y_2) dy_2 \\ &= \frac{1}{4\lambda^2 w_1 w_2} \int_{-\infty}^{-\infty} e^{-\left|\frac{y-y_2}{w_1}\right|/\lambda} e^{-|y_2/w_2|/\lambda} dy_2 \\ &= \frac{w_1 e^{-|y|/w_1 \lambda} - w_2 e^{-|y|/w_2 \lambda}}{2\lambda(w_1 - w_2)(w_1 + w_2)} \end{aligned}$$

2D Model (Laplace-Laplace) (continued)



- Energy Function

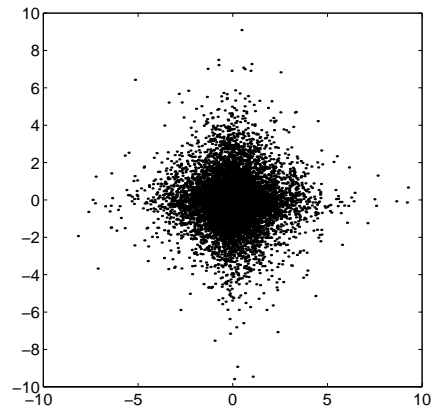
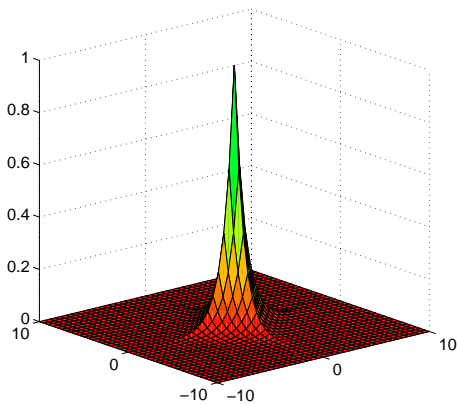
$$R_{\text{BCM}} = \lambda^3 \frac{w_1^5 - w_2^5}{w_1^2 - w_2^2} - \frac{\lambda^4}{4} (w_1^2 + w_2^2)^2$$

$$K_2 = 9\lambda^4 w_1^4 + 6\lambda^4 w_1^2 w_2^2 + \lambda^4 w_2^4 \quad (\|\mathbf{w}^2\| = 1)$$

- Fixed Points

$$\mathbf{w}_{\text{BCM}} = \begin{pmatrix} 0 \\ 3/\lambda \end{pmatrix}, \begin{pmatrix} 3/\lambda \\ 0 \end{pmatrix}$$

$$\mathbf{w}_{K_2} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$



2D Model with 2 Eyes: Normal Rearing

- both eyes see the same thing: $x_1^l = x_1^r, x_2^l = x_2^r$

$$\begin{aligned}y &= w_1 x_1^l + w_2 x_2^l + w_3 x_1^r + w_4 x_2^r \\ &= (w_1 + w_3)x_1 + (w_2 + w_4)x_2 \\ &= w_1^{\text{eff}} x_1 + w_2^{\text{eff}} x_2\end{aligned}$$

- solutions found earlier can be used for w_i^{eff}

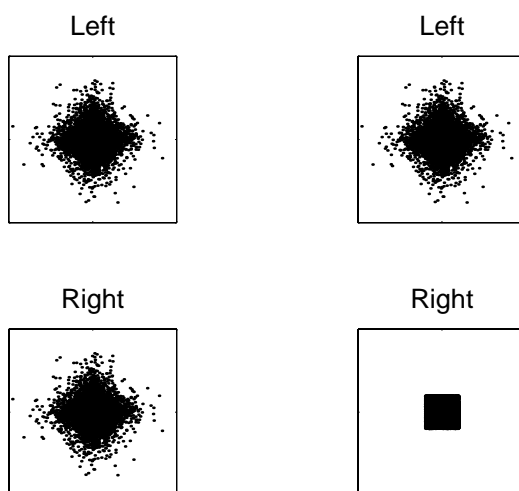
$$\mathbf{w}_{\text{eff}} = \begin{pmatrix} \pm 3/\lambda \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \pm 3/\lambda \end{pmatrix}$$

- since y is shared by both eyes, and \mathbf{x} is the same for both eyes, the only difference is their initial conditions

$$\begin{aligned}\dot{\mathbf{w}}^l - \dot{\mathbf{w}}^r &= 0 \\ \mathbf{w}_0^l &\approx \mathbf{w}_0^r \approx \mathbf{0} \\ \mathbf{w} &= \begin{pmatrix} \pm 3/2\lambda \\ 0 \\ \pm 3/2\lambda \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \pm 3/2\lambda \\ 0 \\ \pm 3/2\lambda \end{pmatrix}\end{aligned}$$

2D Model with 2 Eyes: Monocular Deprivation

- for deprivation, we change the distribution of inputs to one eye

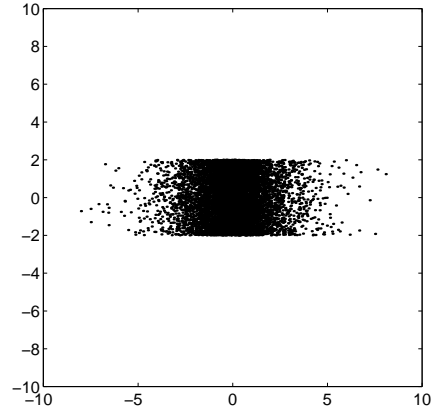
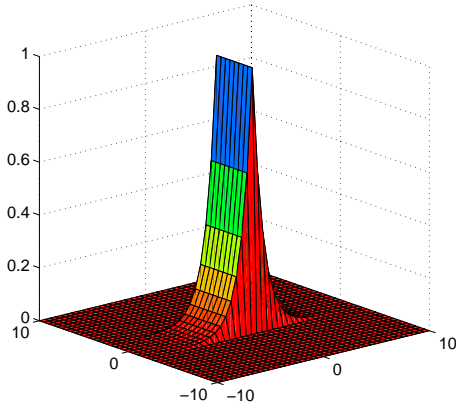


- fixed points only have 1 non-zero value per eye, and inputs are independent, the problem reduces to a 2D problem with initial value

$$\mathbf{w}_0^{\text{eff}} = \begin{pmatrix} \pm 3/2\lambda \\ \pm 3/2\lambda \end{pmatrix}$$

and a 2D environment with one Laplace input, and one Uniform input

2D Model (Laplace-Uniform)



- Energy Functions are messy and non-intuitive
- w_1 represents open eye (laplace numbers), w_2 represents closed eye (uniform numbers)
- Both BCM and Kurtosis have fixed points, for small noise, with $w_2 = 0$
- Opposite dependence on the noise variance: high kurtosis beats low kurtosis
- **BCM**, with low noise approximation

$$\frac{dw_2}{dt} \approx -\frac{3}{2}a^2w_2 + \frac{1}{8}a^3w_2^2$$

- **Kurtosis**, at the beginning

$$\frac{dw_2}{dt} \approx -(12\lambda^4 - 3\lambda^2) + \frac{1}{60}a^4$$

Conclusions

- Neurons modify in order to find significant structure in the data
- Seeking structure can be thought of as maximizing an energy function, though the calculations need to be *local*
- Deprivation can be seen as competing distributions
- Analysis in a simple environment can uncover some hidden differences between different learning rules
- Hopefully, understanding what is important for neurons in this system, will help us understand the behavior of other neural systems