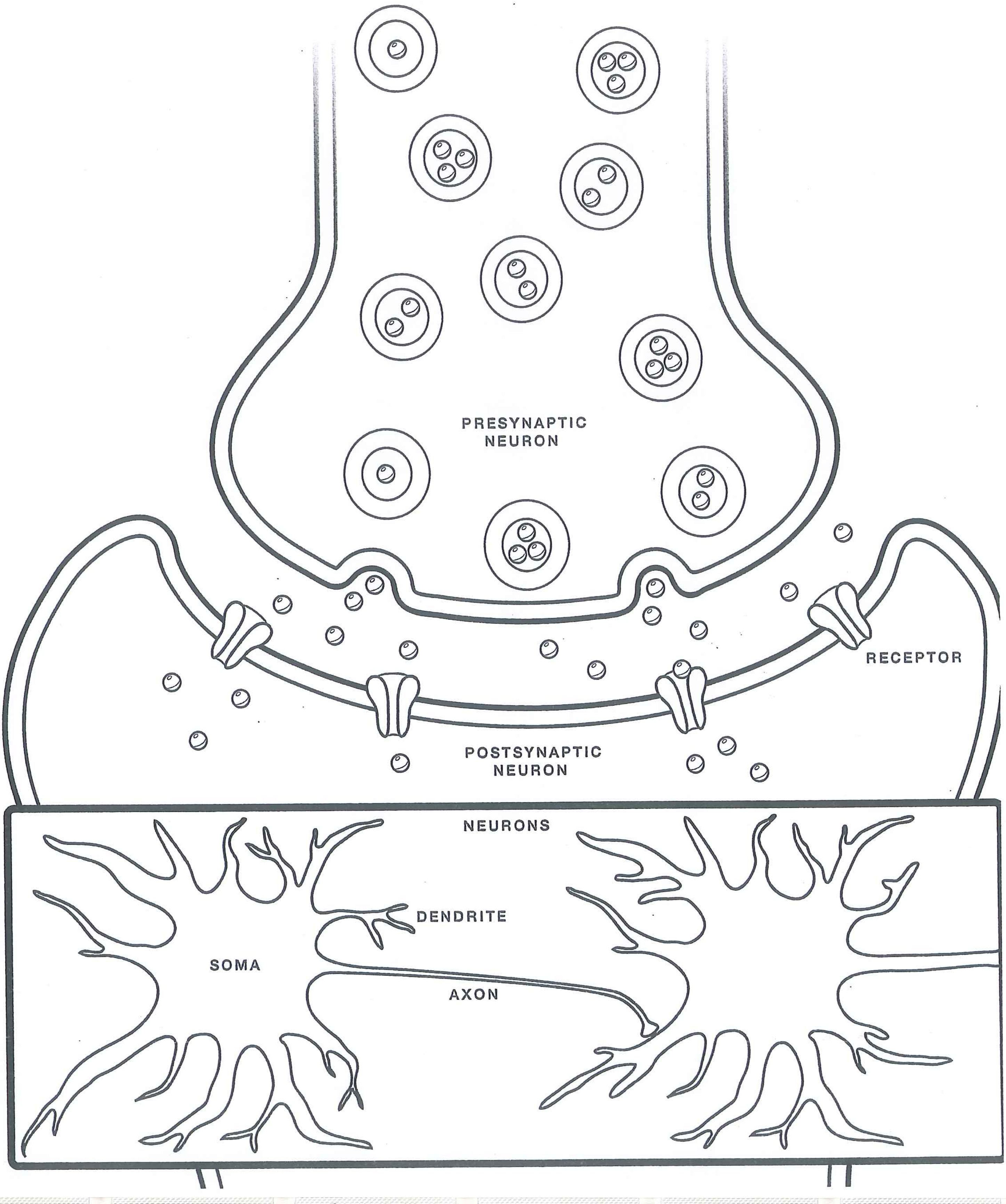
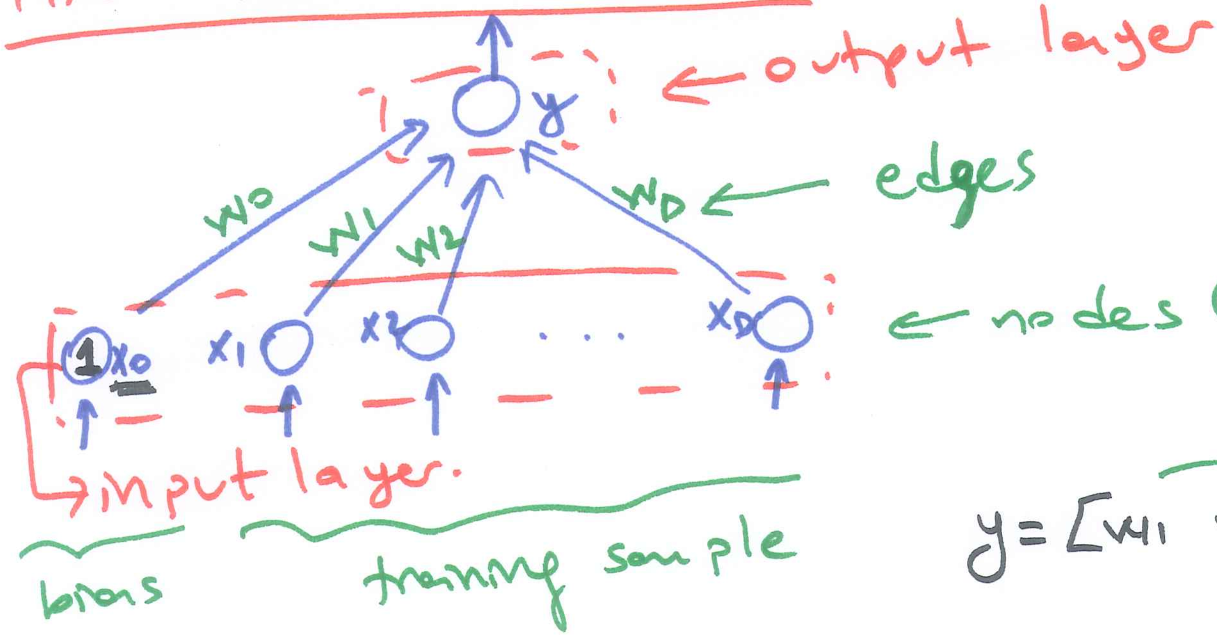
There are 100 billion neurons in your brain.

Neurons communicate with one another at synapses where two cells exchange signals. Here, the postsynaptic neuron takes up molecules released by the presynaptic neuron. Synapses form circuits between cells and are important for learning new things and remembering what you've learned.



THE PERCEPTRON



$$x \in \mathbb{R}^D$$

$$y = s\left[\sum_{d=1}^{D} w_d \cdot x_d + w_0\right]$$

$$y = s\left[\sum_{d=0}^{D} w_d \cdot x_d\right]$$

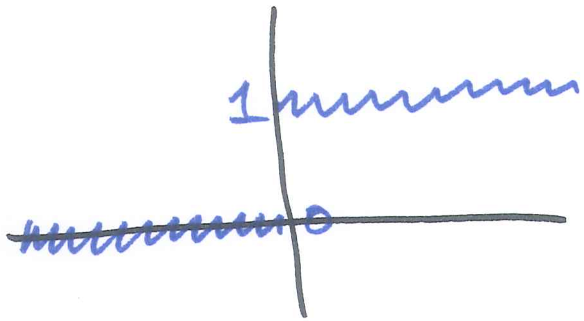
$$y = \overbrace{[w_1 \ w_2 \ \dots\dots \ w_D]}^{w^T} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_D \end{bmatrix}}_{x} + w_0$$

$$y = [w_0 \ w_1 \ w_2 \ \dots\dots \ w_D] \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_D \end{bmatrix}$$

threshold function (activation function)

$$s(a) = \begin{cases} 1 & \text{if } a > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$s(w^T \cdot x) = \begin{cases} 1 & \text{if } w^T x > 0 \\ 0 & \text{otherwise} \end{cases}$$

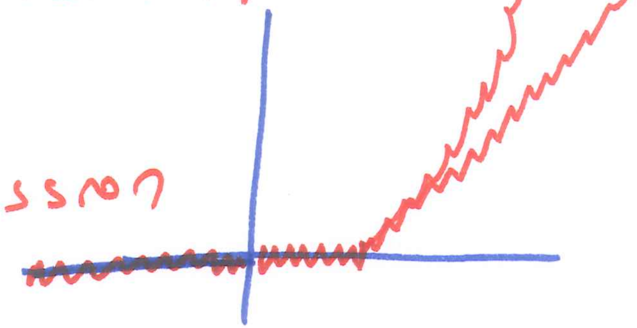


$$y = \text{sigmoid}(w^T x) = \frac{1}{1+\exp[-w^T.x]} \Rightarrow \text{binary classification}$$

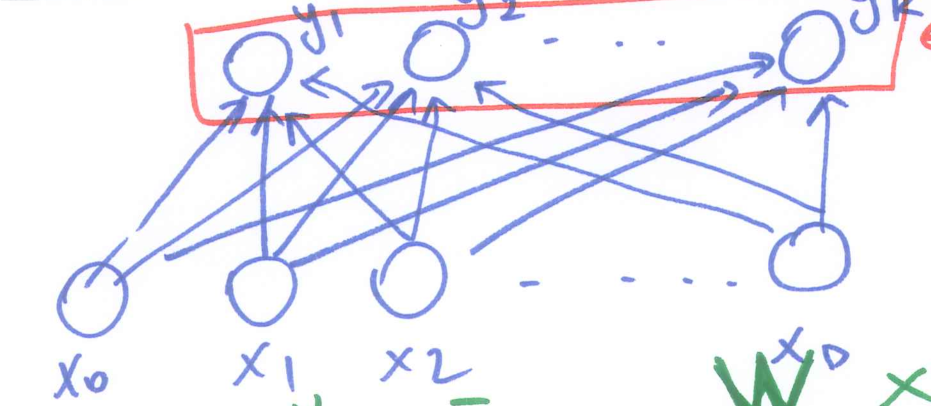
w^T → (D+1) dimensional.

x → (D+1) dimensional.

$$s(a) = a \qquad y = s(w^T.x) = w^T.x \Rightarrow \text{regression}$$



MULTICLASS CLASSIFICATION



$$y_c = s\left[\sum_{d=1}^{D} w_{cd}.x_d + w_{c0}\right] = \boxed{w_c^T.x}$$

$w_c \Rightarrow (D+1)$-

$x \Rightarrow (D+1)$-dimensional

$$y = W x$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_K \end{bmatrix}_{K\times 1} = \begin{bmatrix} w_{10} & w_{11} & w_{12} & \cdots & w_{1D} \\ w_{20} & w_{21} & w_{22} & \cdots & w_{2D} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{K0} & w_{K1} & w_{K2} & \cdots & w_{KD} \end{bmatrix}_{K\times(D+1)} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_D \end{bmatrix}_{(D+1)\times 1}$$

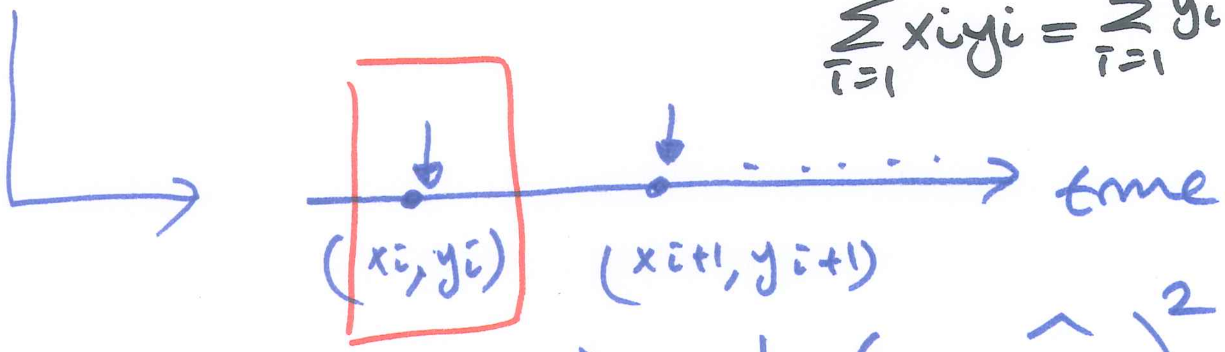
$$y_1 = w_{10}.x_0 + w_{11}.x_1 + \ldots + w_{1D}.x_D$$

$$y_c = \frac{\exp(w_c^T.x)}{\sum_{k=1}^{K} \exp(w_k^T.x)}$$

$$\boxed{\text{choose } \hat{y}^* = \arg\max_{c} y_c}$$ ⇐ decision rule.

Online Learning versus Batch Learning

$$\sum_{i=1}^{N} x_i y_i = \sum_{i=1}^{N} y_i . x_i$$



$$(w^T . x_i)^2 = \cancel{(w^T)^2 . x_i^2}$$

$$w^T . x_i x_i^T . w = x_i^T . w w^T . x_i$$

$$\text{Error}(\underline{\underline{w}} \mid x_i, y_i) = \frac{1}{2} (y_i - \hat{y}_i)^2 \quad \hat{y}_i \rightarrow g(w^T . x_i) = w^T . x_i$$

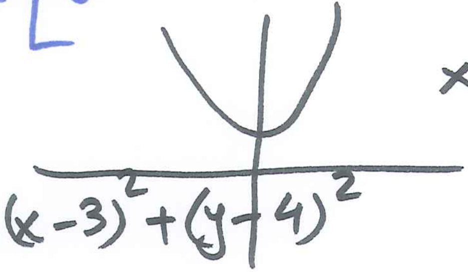
$$= \frac{1}{2} (y_i - w^T . x_i)^2 = \frac{1}{2} \left[y_i . y_i - 2 . y_i \boxed{w^T} x_i + \underbrace{\boxed{w^T}}_{1 \times (D+1)} \underbrace{x_i}_{(D+1) \times 1} \underbrace{x_i^T}_{1 \times (D+1)} . \underbrace{\boxed{w}}_{(D+1) \times 1} \right]$$

$$= \frac{1}{2} \left[y_i^2 - 2 . y_i w^T . x_i + w^T . x_i x_i^T . w \right]$$

$$\frac{\partial x^2}{\partial x} = \underline{\underline{2x}}$$

$$\frac{\partial x x}{\partial x} = x + x = \underline{\underline{2x}}$$

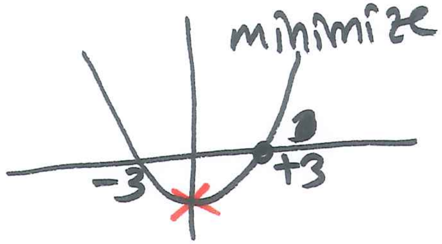
$$\frac{\partial w^T . w}{\partial w} = 2 . w \qquad \frac{\partial a^T . w}{\partial w} = a$$



$x^2 - 9$

$z = (x-3)^2 + (y-4)^2$

minimize $x^2 - 9$



$$\frac{\partial \, Error}{\partial w} = \frac{\partial \left[\frac{1}{2} \left(y_i^2 - 2 y_i w^T . x_i + w^T . x_i x_i^T . w \right) \right]}{\partial w}$$

$$= \cancel{\frac{\partial \frac{1}{2} y_i^2}{\partial w}} \xrightarrow{} 0 \; - \underbrace{\frac{\partial \, y_i w^T . x_i}{\partial w}}_{-\; y_i . x_i} + \underbrace{\frac{\partial \frac{1}{2} \overbrace{w^T . x_i}^{f(w)} \overbrace{x_i^T . w}^{g(w)}}{\partial w}}_{x_i . \hat{y_i}}$$

$$\frac{\partial \frac{1}{2} w^T . x_i . x_i^T . w}{\partial w} = \frac{1}{2} \frac{\partial \, w^T . x_i}{\partial w} . x_i^T . w + \frac{1}{2} \frac{\partial \, x_i^T . w}{\partial w} . w^T . x_i$$

$$= \frac{1}{2} . \frac{x_i \underbrace{x_i^T . w}_{\hat{y_i}}}{} + \frac{1}{2} . \frac{. x_i \underbrace{w^T . x_i}_{\hat{y_i}}}{}$$

$$= \frac{1}{2} . x_i . \hat{y_i} + \frac{1}{2} . x_i . \hat{y_i} = x_i . \hat{y_i}$$

$$\frac{\partial \, Error}{\partial w} = \left(\hat{y_i} - y_i \right) . x_i \qquad \Delta w = - \eta . \frac{\partial \, Error}{\partial w}$$

$$\Delta w = \boxed{\eta \left(y_i - \hat{y_i} \right) . x_i}$$