The Backpropagation Algorithm

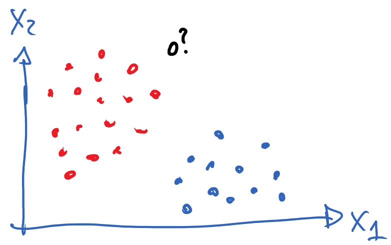
Thursday, September 26, 2019 1:59 PM

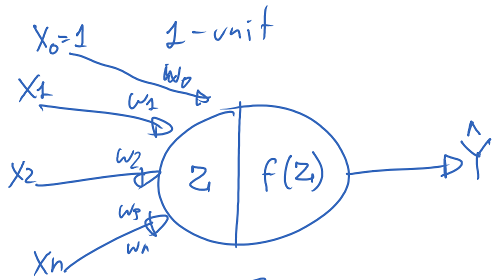
Classification:

Data Examples $(X_1, X_2 \dots X_n)\ Y$ where Y is boolean

Program to learn from examples in order to classify new inputs data.

Example



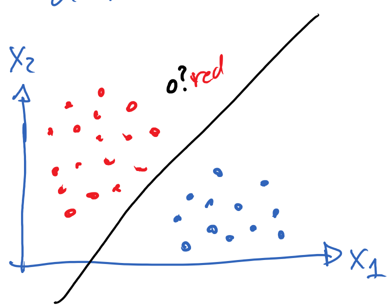


$$Z = \sum_i x_i w_i$$

$$\hat{Y} = f(z)$$

activation function.

Learn w's to partition data:

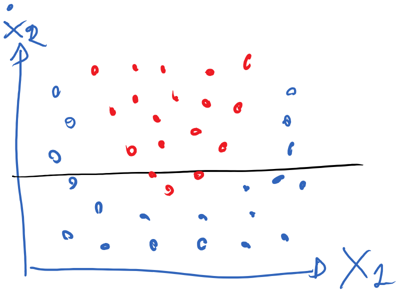


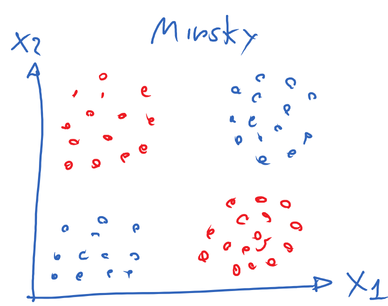
$$f(z) = \begin{cases} 1 & z > 0 \\ 0 & z \leq 0 \end{cases}$$ "perceptron" Rosenblatt.

$$f(z) = sig(z) = \frac{1}{1+e^{z}}$$ sigmoid.

Problems:

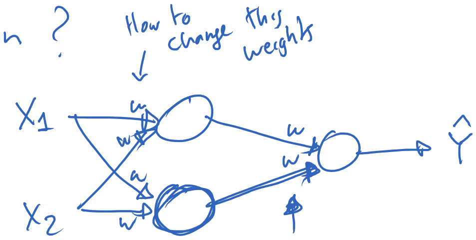
not nicely partitioned





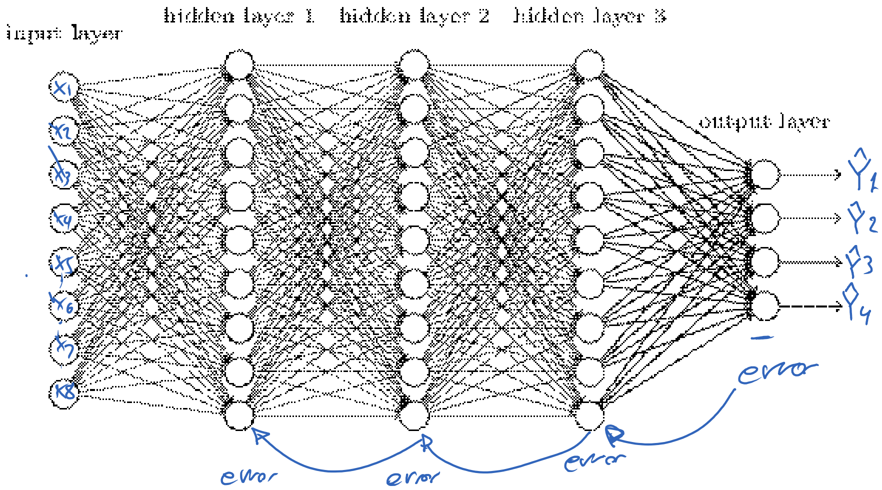
X-or problem.

Solution ?



Back Propagation Algorithm.

- P. Werbos
- D. Rumelhart
- G. Hinton
- R. Williams.



100's of layers
100's of millions of weights

- non-biological.
- non-circuitry

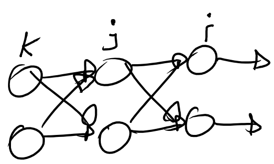
Back-Propagation:-

1.- Prediction.- "feed-forward" step.
given inputs compute value of output.

2.- Back-Propagation
go backwards, layer by layer, updating weights.
- How much to change each weight?
  - in proportion to its contribution to error
- We need to compute "error-term" for each unit.

3.- Repeat 1 and 2 until (hopefully) error is acceptable
....
4.- Profit.

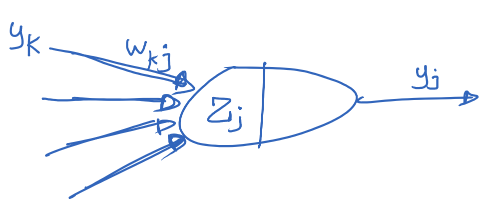
Pre:

- 

- $sig'(z) = sig(z)(1 - sig(z))$

Prediction:

Imagine a unit in layer j



$$z_j = \sum_k w_{kj}\, y_K$$

$$y_j = sig(z_j)$$

- if j is the output layer:
  $y_j = \hat{Y}_j$
- if j is next to the input layer
  $y_j = X_j$

• if j is next to the input layer

$$y_k = X_k$$

Back Propagation:–

chain Rule
$$\frac{\partial f(g(x))}{\partial x} = \frac{\partial f(g(x))}{\partial g(x)} \cdot \frac{\partial g(x)}{\partial x}$$

• output layer:
- suppose layer j is the output layer.

$$\text{error} = \sum_j \Big(\underset{\text{data}}{Y_j} - \underset{\text{output.}}{\hat{Y}_j}\Big)^2 \qquad y_j = \hat{Y}_j$$

$$\text{update } W_{kj} = -\underset{\text{Learning rate}}{\eta} \frac{\partial\, \text{error}}{\partial W_{kj}}$$

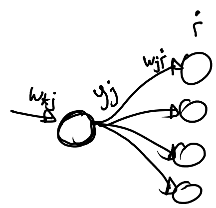
$$= -\eta \cdot \frac{\partial\, \text{error}}{\partial y_j} \cdot \frac{\partial y_j}{\partial z_j} \cdot \frac{\partial z_j}{\partial W_{kj}}$$

$$= -\eta \cdot \big(-2 (Y_j - y_j)\big) \cdot y_j \cdot (1 - y_j)\, y_K$$

$$\text{update } w_{kj} = \eta \cdot (Y_j - y_j) \cdot y_j \cdot (1 - y_j)\, y_K$$

- Inner layer.

suppose j is an inner layer.



$$\text{update } w_{kj} = \eta \frac{\partial\, \text{error}}{w_{kj}}$$

$$= \eta \frac{\partial\, \text{error}}{\partial y_j} \cdot \frac{\partial y_j}{\partial z_j} \cdot \frac{\partial z_j}{\partial w_{kj}}$$

$$\delta_j = \frac{\partial\, \text{error}}{\partial y_j} = \sum_i \frac{\partial\, \text{error}}{\partial y_i} \cdot \frac{\partial y_i}{\partial z_i} \cdot \underset{\cdot\, w_{ji}}{\underset{\parallel}{\frac{\partial z_i}{\partial y_j}}} \qquad \delta_i = \frac{\partial\, \text{error}}{\partial y_i}$$