

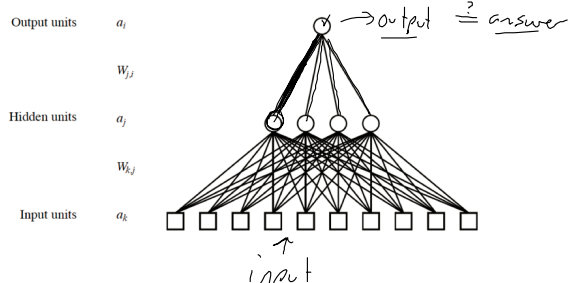
Back Propagation

Friday, November 11, 2016 9:38 AM

Multilayer network learning

• Back-propagation of error

- Each node is responsible for some fraction of the total error based on strength of connections
- Update weights based on the amount of error each node is responsible for
- Search for set of weights that minimize error (gradient descent)



Back propagation

- Search for a set of weights that reduces the error of the output nodes
- Hidden nodes:
 - Each hidden node j is responsible for some of the error in each of the output nodes
 - The greater the weight, the more of the error the hidden node is responsible for
 - The more it should change
- Gradient Descent

loss function

$$\text{loss}(w) = \sum_k (y_k - a_k)^2$$

answ we wanted activation we got

$$\begin{aligned} \frac{\partial}{\partial w} \text{loss}(w) &= \frac{\partial}{\partial w} \sum_k (y_k - a_k)^2 \\ &= \sum_k \frac{\partial}{\partial w} (y_k - a_k)^2 \\ &= \sum_k 2(y_k - a_k) \frac{\partial}{\partial w} (y_k - a_k) \\ &= \sum_k -2(y_k - a_k) \frac{\partial}{\partial w} a_k \leftarrow g'(in_k) \end{aligned}$$

Save anything on the web to OneNote in one click Get OneNote Web Clipper

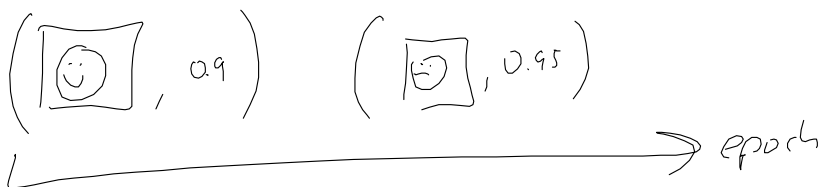
$$\begin{aligned} &= \sum_k -2(y_k - a_k) \frac{\partial}{\partial w} g(in_k) \\ &= \sum_k -2(y_k - a_k) g'(in_k) \frac{\partial}{\partial w} in_k \\ &= \sum_k -2(y_k - a_k) g'(in_k) \frac{\partial}{\partial w} \left(\sum_j w_{jk} d_j \right) \end{aligned}$$

delta improvement Next level down

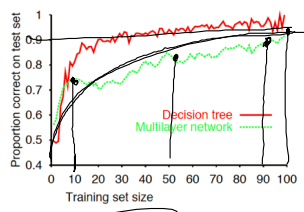
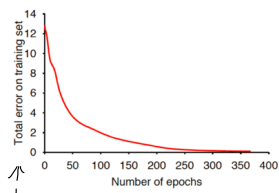
Back propagation

- 1. Initialize weights in network to small random numbers
- 2. Given an example (x, y) , run the network
3. Propagate backwards
 - For each node j in output layer do:
 - $\Delta_j = g'(in_j) \times (y_j - a_j)$ derivative of activation
 - For each layer l from $L-1$ to 1 do:
 - For each node i in layer l do:
 - $\Delta_i = g'(in_i) \sum_j w_{ij} \Delta_j$ step constant to weight update
 - For each weight w_{ij} in network do:

Save anything on the web to OneNote in one click Get OneNote Web Clipper

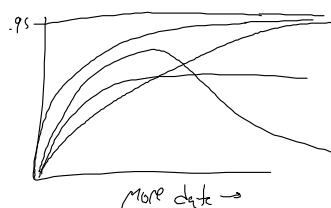


- Error decreases to zero – converges to a perfect fit on training data
- May need to manually tweak network structure to get perfectly optimal



data = testing
training

Happy Graph



Cross-fold validation

train on 10% test on 90%
train on 20% test on 80%
train on 30% test on 70%

Save anything on the web to OneNote in one click Get OneNote Web Clipper

overfitting

Save anything on the web to OneNote in one click

Get OneNote Web Clipper

✕

Save anything on the web to OneNote in one click

Get OneNote Web Clipper

✕

Save anything on the web to OneNote in one click

Get OneNote Web Clipper

✕

Save anything on the web to OneNote in one click

Get OneNote Web Clipper

✕

Save anything on the web to OneNote in one click

Get OneNote Web Clipper

✕

Save anything on the web to OneNote in one click

Get OneNote Web Clipper

✕

Save anything on the web to OneNote in one click

Get OneNote Web Clipper

✕

Save anything on the web to OneNote in one click

Get OneNote Web Clipper

✕

Save anything on the web to OneNote in one click

Get OneNote Web Clipper

X