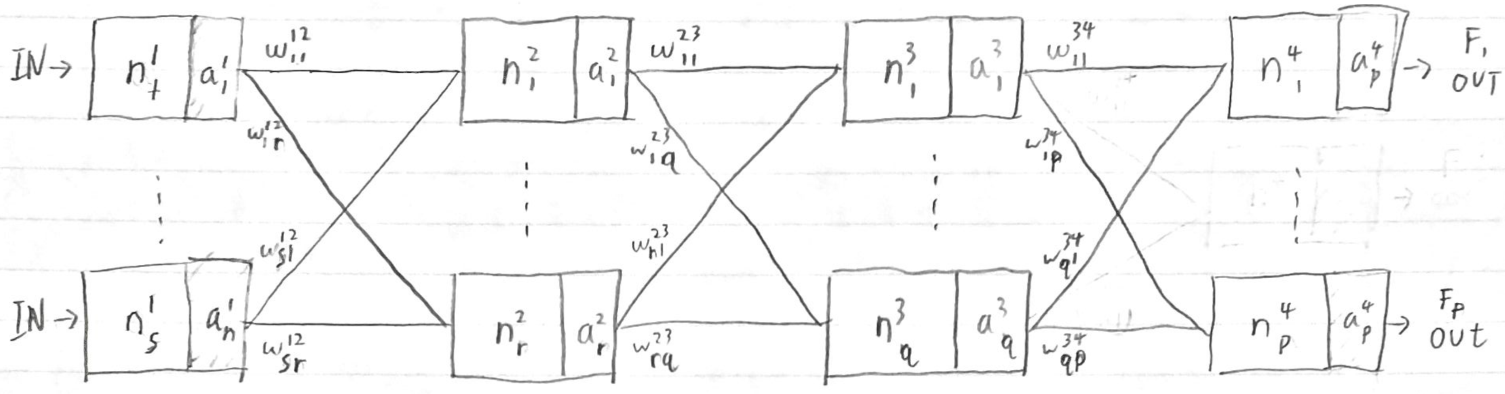
Multilayer Perceptron w/ 3 Weight Layers

3.26.2015 | Kevin Zhang | KZ



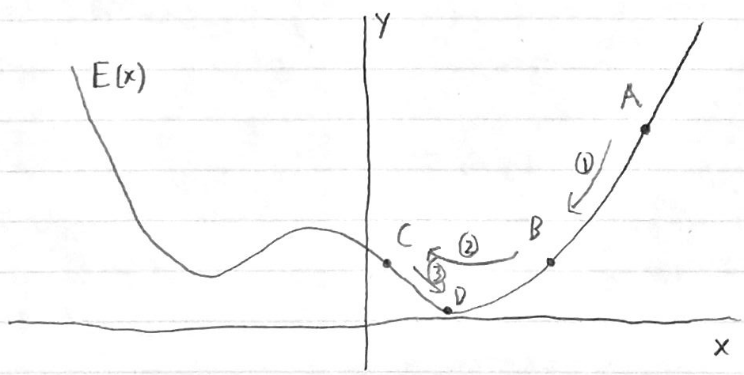
$a^3_q = f(n^3_q)$

$F_p = a^4_p = f(\sum_q a^3_q \cdot w^{34}_{qp})$ where f is the activation function, in our case, the sigmoid

$E = \sum_p \frac{1}{2}(F_p - T_p)^2$ where T_p is the expected value for the pth output

Goal: Minimize the error function E, (and "accidentally" show how backpropagation can be implemented, simply by modifying the structure of the nodes.

How Steepest Gradient Descent Works:



Suppose this is a plot of the error function. Our goal is to minimize the error using steepest gradient descent. Starting at point A, let's go through the steps.

1. What is the derivative of the error function at A? From simply looking at the graph, we can tell $E'(A)$ is positive (because the line slopes upwards).
2. Move in the opposite direction of the direction indicated by the derivative, so $\Delta x = -1 \cdot E'(A)$, and we move to point B.

Adjusting Weights Using Steepest Gradient Descent

$$\frac{dE}{dw^{34}_{qp}} = (F_p - T_p)\cdot \frac{dF_p}{dw^{34}_{qp}} = (F_p - T_p)\cdot f'(n^4_p)\cdot \frac{d(n^4_p)}{d\,w^{34}_{qp}} = (F_p - T_p)\cdot f'(n^4_p)\cdot a^3_q$$

$$\Delta w^{34}_{qp} = -1\cdot \frac{dE}{dw^{34}_{qp}}$$ for all weights in the qp-layer

$\swarrow n^4_p = \sum_q a^3_q \cdot w^{34}_{qp}$ $\swarrow a^3_q = f(n^3_q)$

$$\frac{dE}{dw^{23}_{rq}} = \sum_p (F_p - T_p)\frac{dF_p}{dw^{23}_{rq}} = \sum_p (F_p - T_p)\cdot f'(n^4_p)\cdot \frac{d(n^4_p)}{dw^{23}_{rq}} = \sum_p (F_p - T_p)\cdot f'(n^4_p)\cdot w^{34}_{qp}\cdot \frac{d(a^3_q)}{dw^{23}_{rq}}$$

$$= \sum_p (F_p - T_p)\cdot f'(n^4_p)\cdot w^{34}_{qp}\cdot f'(n^3_q)\cdot \frac{d(n^3_q)}{d(w^{23}_{rq})} = \sum_p (F_p - T_p)\cdot f'(n^4_p)\cdot w^{34}_{qp}\cdot f'(n^3_q)\cdot a^2_r$$

$$\Delta w^{23}_{rq} = -1\cdot \frac{dE}{dw^{23}_{rq}}$$ for all weights in the rq-layer

$\swarrow n^4_p = \sum_q a^3_q \cdot w^{34}_{qp}$ $\swarrow n^3_q = \sum_r a^2_r \cdot w^{23}_{rq}$

$$\frac{dE}{dw^{12}_{sr}} = \sum_p (F_p - T_p)\cdot \frac{dF_p}{dw^{12}_{sr}} = \sum_p (F_p - T_p)\cdot f'(n^4_p)\cdot \frac{d(n^4_p)}{dw^{12}_{sr}} = \sum_p (F_p - T_p)\cdot f'(n^4_p)\cdot\left(\sum_q w^{34}_{qp}\cdot f'(n^3_q)\cdot \frac{dn^3_q}{dw^{12}_{sp}}\right)$$

$$= \sum_p (F_p - T_p)\cdot f'(n^4_p)\cdot\left(\sum_q w^{34}_{qp}\cdot f'(n^3_q)\cdot\left(w^{23}_{rq}\cdot \frac{d(a^2_r)}{dw^{12}_{sr}}\right)\right) = \sum_p (F_p - T_p)\cdot f'(n^4_p)\cdot\left(\sum_q w^{34}_{qp}\cdot f'(n^3_q)\cdot\left(w^{23}_{rq}\cdot f'(n^2_r)\cdot a^1_s\right)\right)$$

$$\Delta w^{12}_{sr} = -1\cdot \frac{dE}{dw^{12}_{sr}}$$ for all weights in the sr-layer

---

Backpropagation / Efficient Gradient Descent

$$\Delta w^{34}_{qp} = -1\cdot \underline{(F_p - T_p)\cdot f'(n^4_p)}\cdot a^3_q$$

$$\Delta w^{23}_{rq} = -1\cdot \sum_p \underline{(F_p - T_p)\cdot f'(n^4_p)}\cdot \underline{w^{34}_{qp}\cdot f'(n^3_q)}\cdot a^2_r$$

$$\Delta w^{12}_{sr} = -1\cdot \sum_p \underline{(F_p - T_p)\cdot f'(n^4_p)}\cdot\left(\sum_q \underline{w^{34}_{qp}\cdot f'(n^3_q)}\cdot\left(\underline{w^{23}_{rq}\cdot f'(n^2_r)}\cdot a^1_s\right)\right)$$

<u>Goal:</u> Avoid recalculating the underlined values by carefully arranging the nested loops and storing the intermediate values in arrays (memoization).