Class Notes: DL

Shivendra

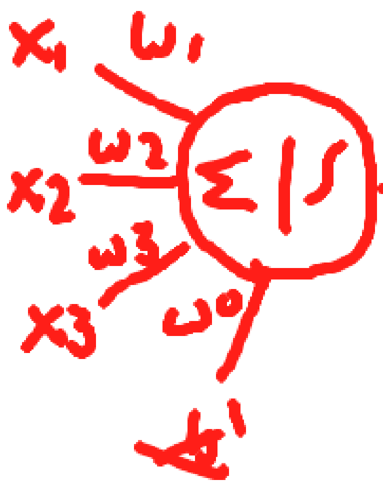
Data

- Data sufficiency
- Data Quality
- Some examples of evaluating information inside data

Basics of Deep Learning

- Deep learning emulation of human brain
- Deep Learning – layered architecture, hierarchical extraction of features
- Deep learning – every unit is capable of linear and non linear processing.

Training a neuron



$$\hat{y} = NL\left(\sum w_i x_i + b\right) = NL\left(\sum w_i x_i\right)$$

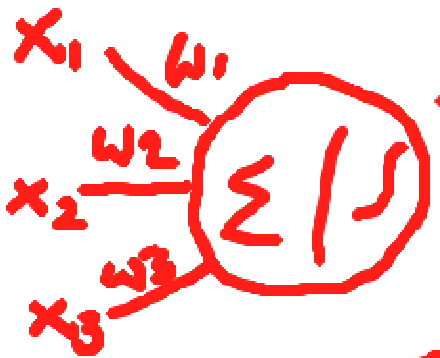
$$SP = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_3$$
$$= \sum w_i x_i$$

$$\hat{y} = NL\left(\sum w_i x_i\right) = Sig\left(\sum w_i x_i\right)$$

↑ Model's Pred

$$E = \underbrace{\sum (y - \hat{y})^2}_{\text{Minimize this Error}} \leftarrow SSE$$

Minimizing the Error



$\hat{y} = Sig(\sum w_i x_i)$ $\quad g = Sig$

$$E = \sum (y - \hat{y})^2 = \sum (y - \underline{Sig(\sum w_i x_i)})^2$$

$$= \sum (y - g(\sum w_i x_i))^2$$

Minimize this Error

$E \to 0$

Car at a vel v, apply brake (multiple)

$\boxed{-\frac{dv}{dt}}$ $\quad E \to 0 \quad$ $\boxed{-\eta \frac{\partial E}{\partial w_i} = \Delta w_i}$

Scaling

Training and Gradient Descent

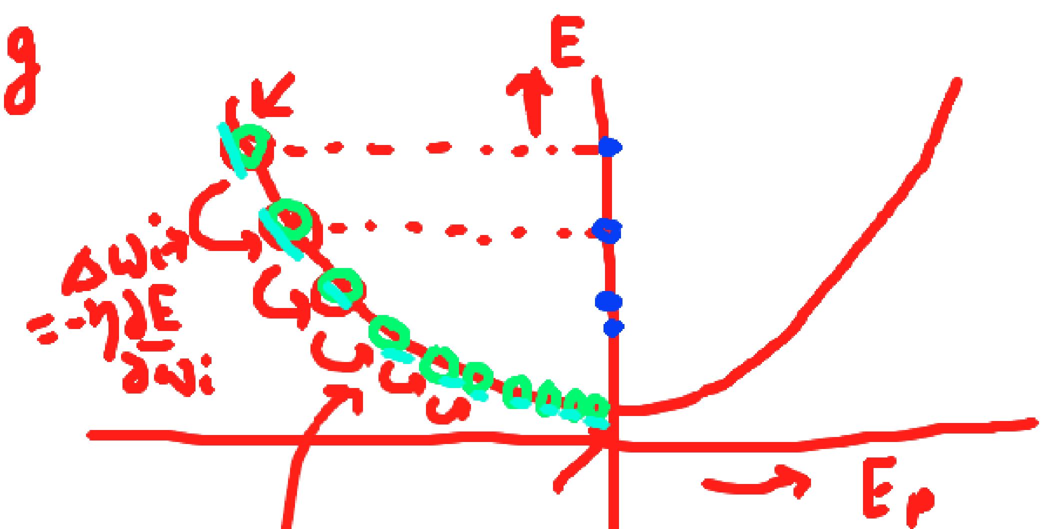
$$\boxed{\Delta w_i = -\eta \frac{\partial E}{\partial w_i}}$$

Minimize the Error

Multiple Times

Scaling Learning Rate

$$E = \sum (y - \hat{y})^2 = \sum \left(y - g\left(\sum w_i x_i\right)\right)^2$$



E↓ Grad↓

Gradient Descent

Derivative of sigmoid

$$g(x) = \mathrm{Sig}(x) = \frac{1}{1+e^{-x}} \qquad \frac{\partial}{\partial x}\left(\mathrm{Sig}(x)\right) = \frac{\partial}{\partial x}\left(\frac{1}{1+e^{-x}}\right)$$

$$= \frac{\partial}{\partial x}\underbrace{(1+e^{-x})}_{u}{}^{-1} \qquad \frac{\partial}{\partial x} u^{-1} = -1 \cdot u^{-1-1} \cdot \frac{\partial u}{\partial x}$$

$$= -1 \cdot (1+e^{-x})^{-2} \frac{\partial}{\partial x}(1+e^{-x}) = \frac{-1}{(1+e^{-x})^2} \cdot (0 + e^{-x} \cdot -1)$$

$$= \frac{e^{-x}}{(1+e^{-x})^2} = \frac{1}{(1+e^{-x})} \cdot \frac{e^{-x}}{(1+e^{-x})} = g(x) \cdot \frac{e^{-x}+1-1}{1+e^{-x}}$$

$$= g(x) \cdot \left(1 - \frac{1}{1+e^{-x}}\right) = g(x) \cdot (1 - g(x))$$

Derivative of Error

$$\boxed{\Delta w_i = -\eta \frac{\partial E}{\partial w_i}}$$ ← Multiple times (Epochs)

$$\frac{\partial E}{\partial w_i} = \frac{\partial}{\partial w_i}\left(\sum \Big(y - \overbrace{g\Big(\sum w_i x_i\Big)}^{\hat{y}}\Big)^2\right)$$

$$= \frac{\partial}{\partial w_i}\left(\sum \Big(\underbrace{y - g\Big(\sum w_i x_i\Big)}_{u}\Big)^2\right) \qquad \frac{\partial}{\partial w_i} u^2 = 2 \cdot u^{2-1} \cdot \frac{\partial u}{\partial w_i}$$

$$= \sum 2 \cdot \Big(y - g\Big(\sum w_i x_i\Big)\Big) \cdot \frac{\partial}{\partial w_i}\Big(y - g\Big(\sum w_i x_i\Big)\Big)$$

$$= 2 \sum \Big(y - g\Big(\sum w_i x_i\Big)\Big) \cdot \left(0 - \frac{\partial}{\partial w_i} g\Big(\sum w_i x_i\Big)\right)$$

Derivative of Error, dependencies

$$\frac{\partial E}{\partial w_i} = 2 \sum \left(y - g\left(\sum w_i x_i\right)\right) \cdot \left(0 - \frac{\partial}{\partial w_i} g\left(\sum w_i x_i\right)\right)$$

$$= 2 \sum \left(y - g\left(\sum w_i x_i\right)\right) \cdot \left(- g(w_i x_i) \cdot \left(1 - g\left(\sum w_i x_i\right)\right)\right) \cdot \frac{\partial}{\partial w_i}\left(\sum w_i x_i\right)$$

$$= -2 \sum \left(y - g\left(\sum w_i x_i\right)\right) \cdot g\left(\sum w_i x_i\right)\left(1 - g\left(\sum w_i x_i\right)\right) \cdot x_i$$

$$\boxed{\frac{\partial E}{\partial w_i} = -2 \sum \underbrace{\left(y - g\left(\sum w_i x_i\right)\right)}_{\text{Dev from True Val}} \cdot \underbrace{g\left(\sum w_i x_i\right) \cdot \left(1 - g\left(\sum w_i x_i\right)\right)}_{\text{Der (Act Fn)}} \cdot \underbrace{x_i}_{\text{Input}}}$$

→ Dev from True Val

→ Der (Act Fn)

→ Input

$$\boxed{\Delta w_i = -\eta \frac{\partial E}{\partial w_i}}$$

$\frac{\partial E}{\partial w_i} \Rightarrow 0$ Trg will stop