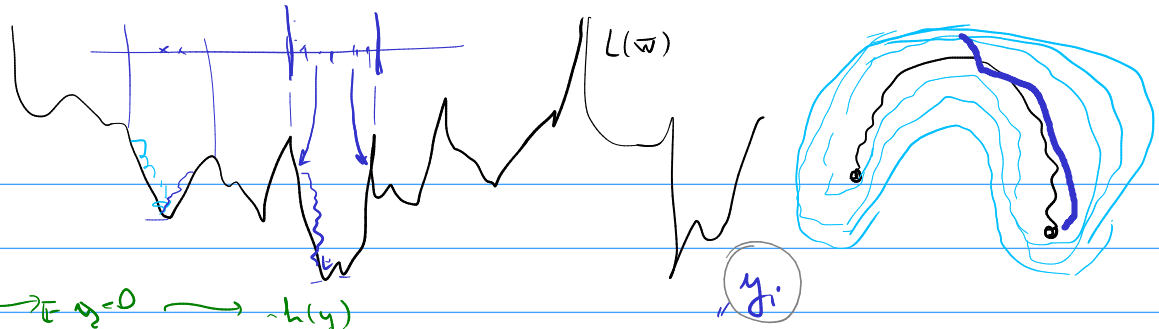
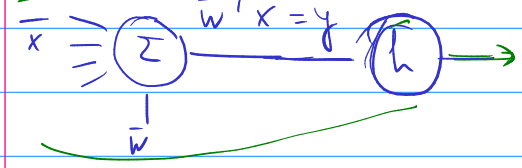
Weight init



Xavier init

He init

$E[x_i] = 0$



$\bar{w}^T \bar{x} = y$

y_i

$y = \bar{w}^T \bar{x} = \sum_i w_i x_i$

independent

$Var(y_i) = Var(w_i x_i) = E[(w_i x_i)^2] - (E[w_i x_i])^2 =$

$E w_i^2 \cdot E x_i^2 \qquad E w_i \cdot E x_i$

$= ((E w_i)^2 + Var\, w_i)((E x_i)^2 + Var\, x_i) - (E w_i)^2 \cdot (E x_i)^2$

$(E w_i)^2 + Var\, w_i$

$w \sim \frac{A}{\sqrt{n}}$

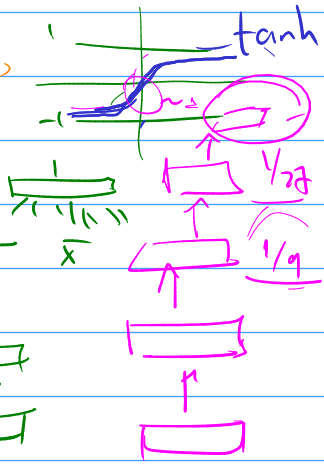
$$Var\, y_i = (Var\, w_i)(Var\, x_i) + (E w_i)^2 \cdot Var\, x_i + (E x_i)^2 \cdot Var\, w_i$$

$E w_i = 0$

Symmetric activ. function:

$$Var\, y_i = Var\, w_i \cdot Var\, x_i$$

$$Var\, y = Var(\sum y_i) = \sum_i Var\, y_i = n \cdot Var\, w_i \cdot Var\, x_i$$



$$Var\, y = n \cdot Var\, w_i \cdot Var\, x_i$$

Neural Networks: tricks of the trade

$w_i \sim Unif\left(\left[-\frac{\sqrt{3}}{\sqrt{n}}, \frac{\sqrt{3}}{\sqrt{n}}\right]\right)$

$Unif[a,b]$

$Var = \frac{(b-a)^2}{12}$

$Var\, w_i = \frac{(2/\sqrt{n})^2}{12} = \frac{1}{3n}$

$Var\, w_i = O\left(\frac{1}{n}\right)$

$Var\, y \sim \frac{1}{3} Var\, x_i$

Xavier

Glorot

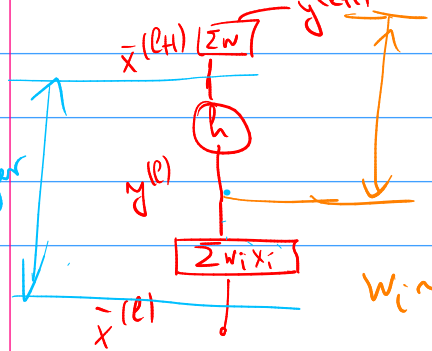
All act. func:

$$Var\, y_i = Var\, w_i \cdot Var\, x_i + (E x_i)^2 \cdot Var\, w_i = Var\, w_i \cdot E[x_i^2]$$

2010

ReLU $f(x) = \max(0, x)$

$E[x^{(l+1)2}] = \frac{1}{2} Var[y^{(l)}]$ ← only for ReLU



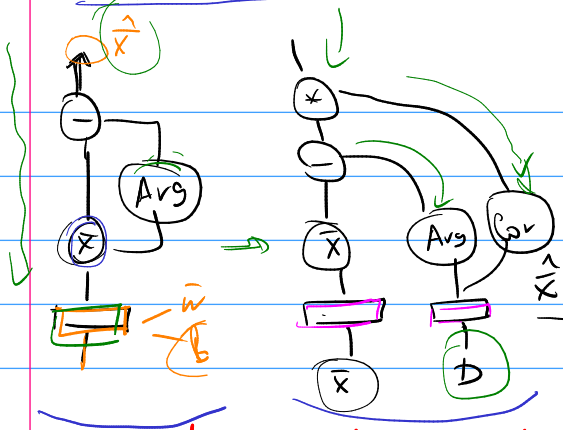
$$Var\, y^{(l+1)} = n_l \cdot \frac{1}{2} \cdot Var[y^{(l)}] \cdot Var\, w_i$$

$$Var\, y^{(l+1)} = \frac{n_l}{2} \cdot Var\, w_i \cdot Var\, y^{(l)}$$

$w_i \sim N(0, \sqrt{\frac{2}{n}})$

$w_i \sim E w_i = 0, \quad Var\, w_i \simeq \frac{2}{n_l} \left[-\frac{\sqrt{6}}{\sqrt{n}}, \frac{\sqrt{6}}{\sqrt{n}}\right]$

Batchnorm



$$\hat{\bar{x}} = \bar{x} - \mathbb{E}[\bar{x}]$$

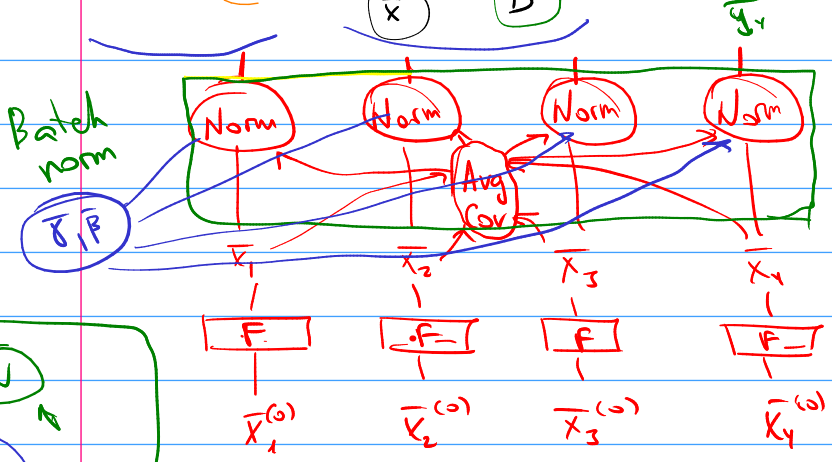
$$\bar{x} = \bar{u} + b$$

$$\bar{x} = \bar{u} + (b + \Delta b) \rightsquigarrow \hat{x} = \bar{x} - \mathbb{E}[\bar{x}] = \bar{u} - \mathbb{E}[\bar{u}]$$

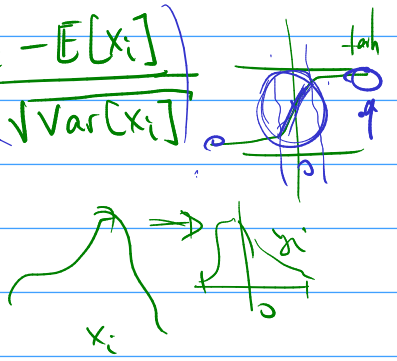
$$\hat{\bar{x}} = Cov_D^{-1}\left(\bar{x} - \mathbb{E}_D[\bar{x}]\right)$$

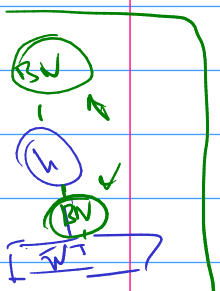
$$\bar{x} = \frac{\bar{x} - \bar{u}}{(.. \; Var x_i \; ..)}$$

$[\mathbb{E} x_i]$



$$y_i = \frac{x_i - \mathbb{E}[x_i]}{\sqrt{Var[x_i]}}$$





$$BN(\bar{x})_i = \gamma_i \cdot \frac{x_i - \mathbb{E} x_i}{\sqrt{Var x_i}} + \beta_i$$

Inc. Keynot Efficient Net DenseNet