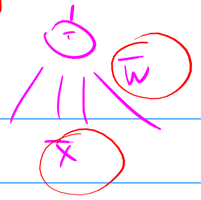
$y = \bar{w}^T \bar{x}$ — h — $h(\bar{w}^T \bar{x})$



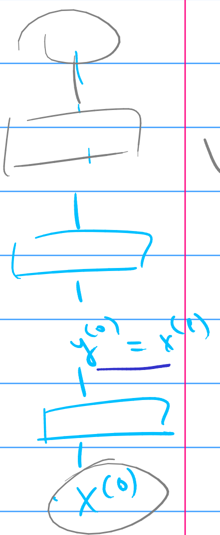
$y = \bar{w}^T \bar{x} = \sum^n w_i x_i = \sum y_i$ ($y_i = w_i x_i$)

$Var(x) = E x^2 - (E x)^2$

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

$= [E x_i]^2 \cancel{Var[w_i]} + [E\, w_i]^2 \cancel{Var[x_i]} +$

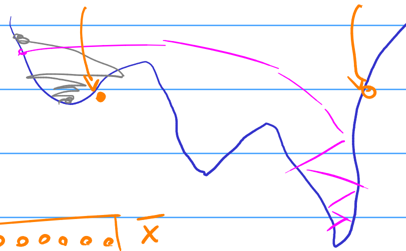
$+ Var[w_i] Var[x_i]$



$Var[y_i] = Var[w_i] Var[x_i]$

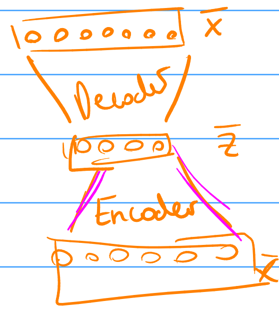
$Var[y] = \boxed{n \cdot Var[w_i]} Var[x_i]$

$Var\, y \sim Var\, x_i$



$Var\, w_i \simeq O\left(\frac{1}{n}\right)$ $1/3n$

$w_i \sim Unif\left[-\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\right]$

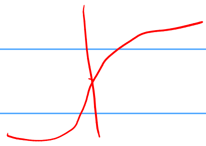


Xavier Glorot $Var[Unif([a,b])] = \frac{(b-a)^2}{12}$

$Var\, y^{(l)} = \frac{1}{3} Var\, x^{(l)}$

Xavier init $w_i \sim U\left[-\sqrt{\frac{3}{n}}, \sqrt{\frac{3}{n}}\right]$

$w_i \sim \mathcal{N}\left(0, \frac{1}{n}\right)$

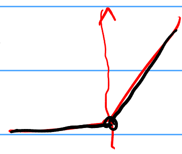


$x^{(l+1)} = h(y^{(l)})$

ReLU

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

$+ Var\, w_i \cdot Var\, x_i^2$



$y^{(l)} = \bar{w}^{(l)T} \bar{x}^{(l)}$

$= Var\, w_i \cdot E[x_i^2]$

Kaiming He

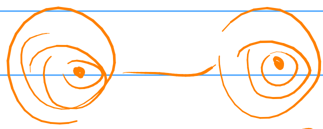
$x^{(l)}$

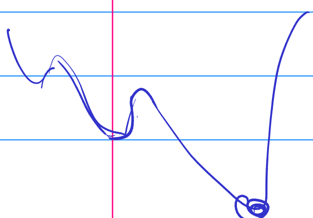
$x^{(l+1)} = \max(0, y^{(l)})$

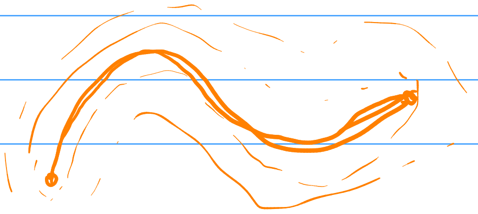
He init

$Var\, y^{(l)} = \frac{n}{2} Var\, w_i^{(l)}\ Var\, y^{(l-1)}$

$y^{(l-1)}$







$$F(\bar{x}) = \mathbb{E}_{q(\bar{y})} f(\bar{x}, \bar{y}) \xrightarrow[\bar{x}]{} \min$$ empirical risk

$$F(\bar{x}) = \frac{1}{N}\sum_{n=1}^{N} f(\bar{x}, d_n) = \underbrace{\mathbb{E}_{n \sim \mathrm{Unif}[1,N]}}_{q(n)} f(\bar{x}, d_n) \xrightarrow[\bar{x}]{} \min$$

SGD

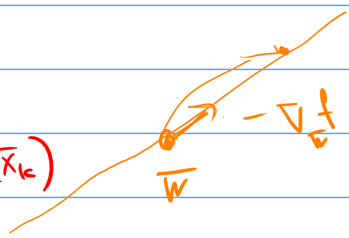
$$\hat{F}(\bar{x}) = \frac{1}{m}\sum_{i}^{m} f(\bar{x}, \bar{y}_i)$$

$$\hat{g}(\bar{x}) = \frac{1}{m}\sum_i \nabla_{\bar{x}} f(\bar{x}, \bar{y}_i)$$

residue

$$\|\bar{x}_k - \bar{x}^*\| \le ?$$

$$\bar{x}_{k+1} = \bar{x}_k - \alpha_k \hat{g}_k$$



$$\mathbb{E}\, \hat{g}_k = \bar{g}_k = \nabla_{\bar{x}} F(\bar{x}_k)$$

$$\|\bar{x}_{k+1} - \bar{x}^*\|^2 = \|\bar{x}_k - \alpha_k \hat{g}_k - \bar{x}^*\|^2 =$$

$$= \|\bar{x}_k - \bar{x}^*\|^2 - 2\alpha_k \hat{g}_k^T(\bar{x}_k - \bar{x}^*) + \alpha_k^2 \|\hat{g}_k\|^2$$

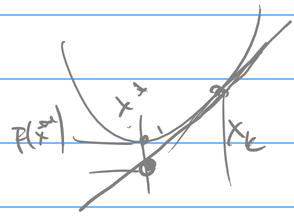
$$\nabla F = \sum_d \nabla_{\bar{w}} f(\bar{w}, d)$$

$$\mathbb{E}_k \|\bar{x}_{k+1} - \bar{x}^*\|^2 = \|\bar{x}_k - \bar{x}^*\|^2 - \underline{2\alpha_k \bar{g}_k^T(\bar{x}_k - \bar{x}^*)} + \alpha_k^2 \mathbb{E}_k \|\hat{g}_k\|^2$$

$$\nabla_{\bar{w}} F_k = \sum_{d \in D_k} \nabla_{\bar{w}} f(\bar{w}, d)$$

$$F(\bar{x}^*) \ge F(\bar{x}_k) - \bar{g}_k^T(\bar{x}_k - \bar{x}^*)$$

$$\alpha_k \bar{g}_k^T(\bar{x}_k - \bar{x}^*) \ge \alpha_k\left(F(\bar{x}_k) - F(\bar{x}^*)\right)$$



$$\alpha_k\left(\mathbb{E} F(\bar{x}_k) - F(\bar{x}^*)\right) \le \frac{1}{2}\mathbb{E}\|\bar{x}_k - \bar{x}^*\|^2 + \frac{1}{2}\alpha_k^2 \mathbb{E}_k\|\hat{g}_k\|^2 - \frac{1}{2}\mathbb{E}_k\|\bar{x}_{k+1} - \bar{x}^*\|^2 \qquad \sum_{i=0}^{k}$$

$$\sum_{i=0}^{k} \alpha_i\left(\mathbb{E} F(\bar{x}_i) - F(\bar{x}^*)\right) \le \frac{1}{2}\|\bar{x}_0 - \bar{x}^*\|^2 + \frac{1}{2}\sum_i \alpha_i^2 \mathbb{E}\|\hat{g}_i\|^2$$

$$\frac{\sum_i \alpha_i \mathbb{E} F(x_i)}{\sum \alpha_i} \ge \mathbb{E} F\left(\frac{\sum_i \alpha_i \bar{x}_i}{\sum \alpha_i}\right)$$

$$\boxed{\mathbb{E} F\left(\frac{\sum_i \alpha_i \bar{x}_i}{\sum \alpha_i}\right) - F(\bar{x}^*)} \le \frac{\sum_i \alpha_i\left(\mathbb{E} F(\bar{x}_i) - F(\bar{x}^*)\right)}{\sum \alpha_i} \le$$

$$\le \frac{\frac{1}{2}\|x_0 - \bar{x}^*\|^2 + \frac{1}{2}\sum_i \alpha_i^2 \mathbb{E}\|\hat{g}\|^2}{\sum_i \alpha_i}$$

$$\|\bar{x}_0 - \bar{x}^*\| \le R$$

$$\mathbb{E}\|\hat{g}_k\|^2 \le G^2$$

$$\mathbb{E} F(\bar{x}_k) - F(\bar{x}^*) \le \frac{R^2 + G^2 \sum_0^k \alpha_i^2}{2\sum_0^k \alpha_i}$$