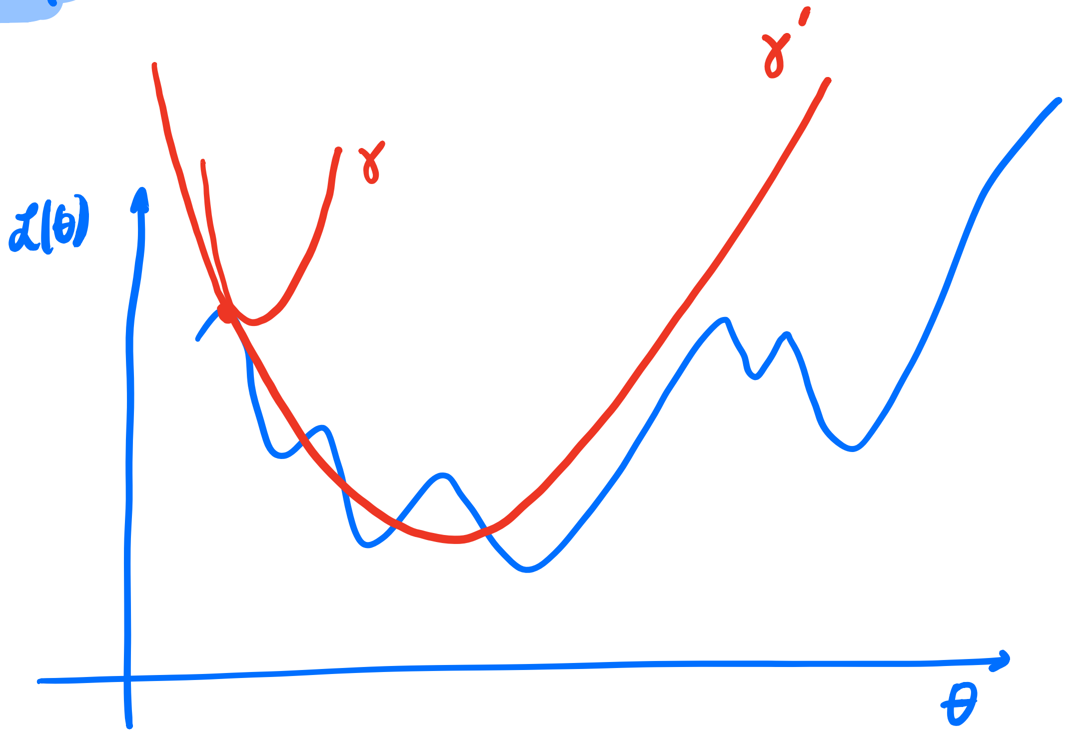
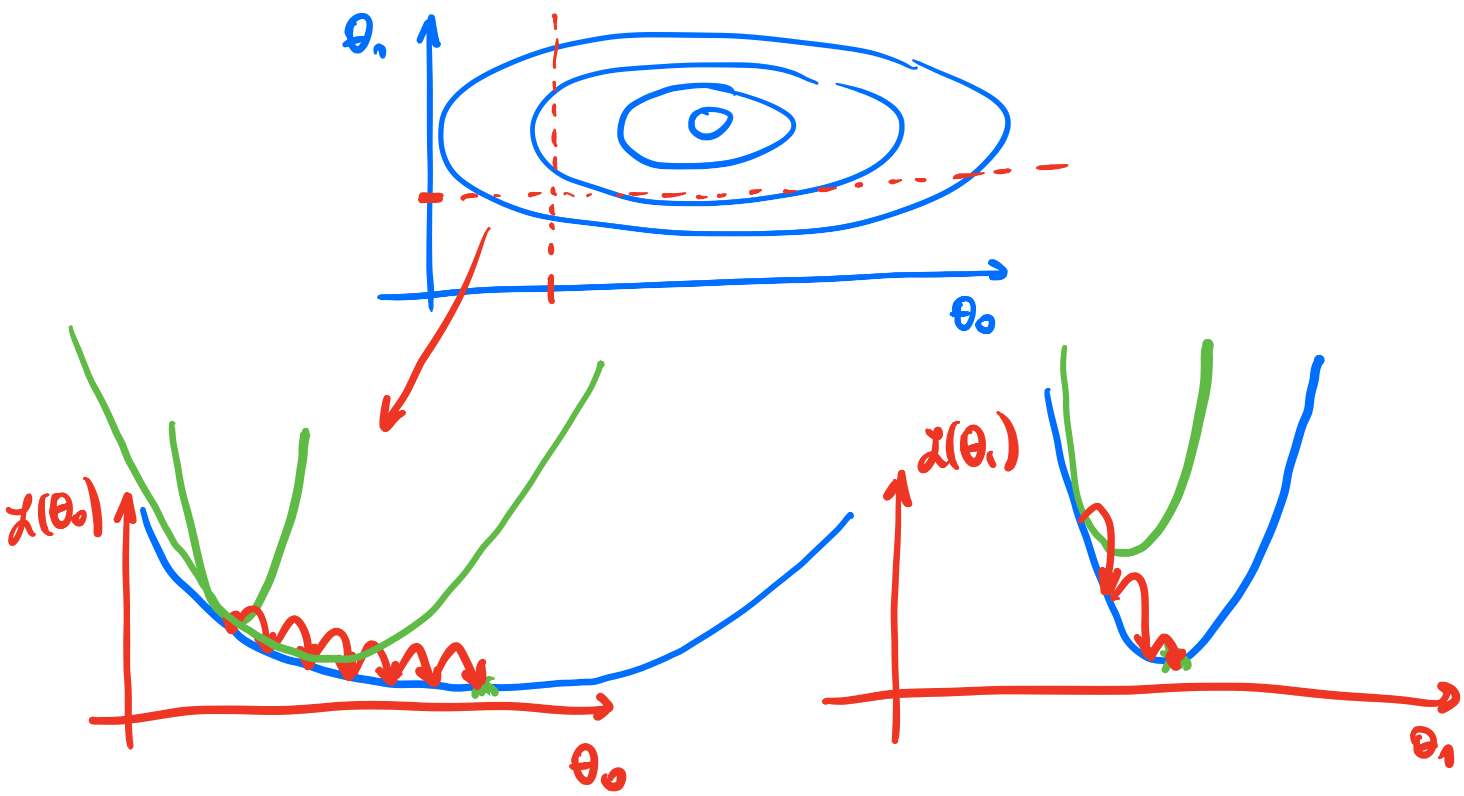
Lecture 4

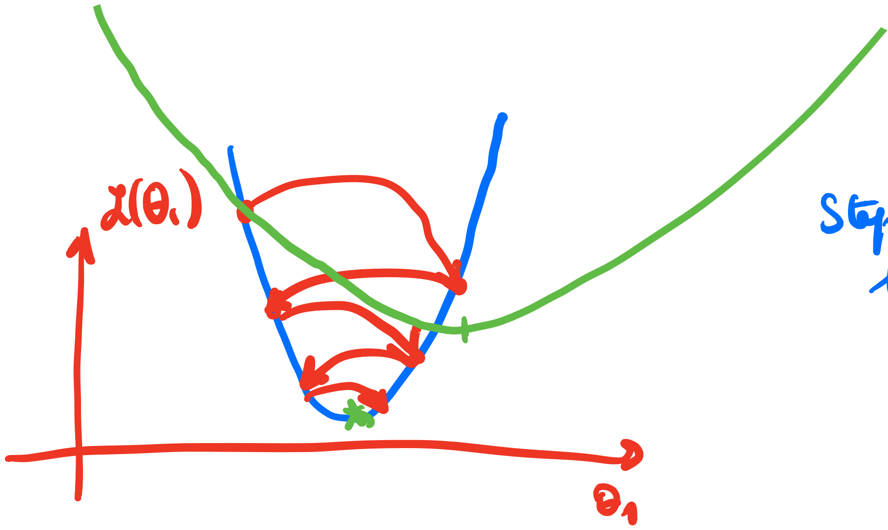
Slide 15



Slide 17



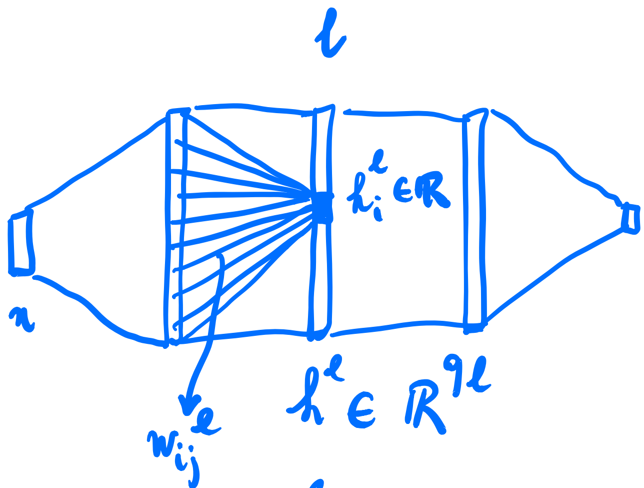
Slide 19



Steps are too large!

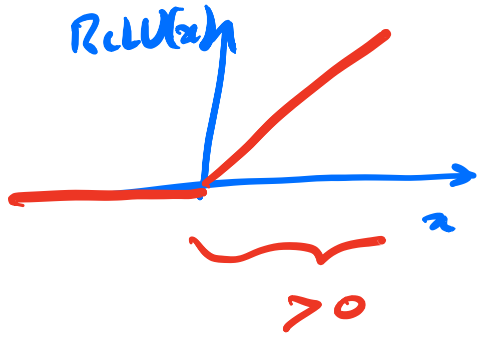
Slide 42

Forward pass:



$$h_i^\ell = \mathrm{ReLU}\left(w_{i:}^T h^{\ell-1} + b_i^\ell\right)$$

$$= \sum_{j=0}^{q_{\ell-1}-1} w_{ij} h_j^{\ell-1}$$



$$\mathbb{V}(A+B) = \mathbb{V}(A) + \mathbb{V}(B) + \mathrm{Cov}(A,B)$$

$$\mathbb{V}(AB) = \mathbb{V}(A)\,\mathbb{V}(B) + \mathbb{V}(A)\,\mathbb{E}(B) + \mathbb{V}(B)\,\mathbb{E}(A)$$

$$\mathbb{V}[h_i^\ell] = \mathbb{V}\left[\sum_{j=0}^{q_{\ell-1}} w_{ij} h_j^{\ell-1}\right]$$

$$= \sum_{j=0}^{q_{\ell-1}} \mathbb{V}[w_{ij}]\,\mathbb{V}[h_j^{\ell-1}]$$

$$\mathbb{V}[h^\ell] = \underbrace{q_{\ell-1}\,\mathbb{V}[w_{ij}]}_{=1}\,\mathbb{V}[h^{\ell-1}]$$

$$\Rightarrow \mathbb{V}[w_{ij}] = \frac{1}{q_{\ell-1}}$$

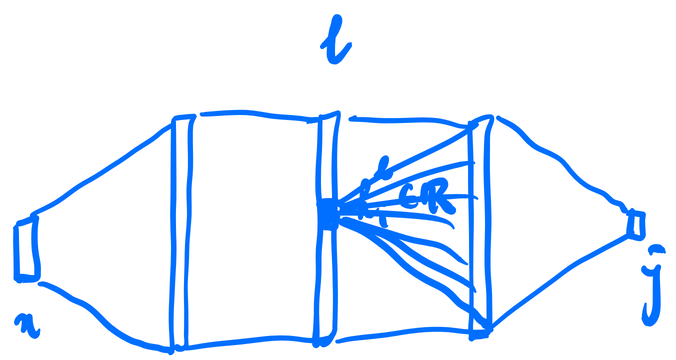
LeCun initialization

$$U[a, b] \longrightarrow V = \frac{(b-a)^2}{12}$$

$$U[-a, a] \longrightarrow V = \frac{4a^2}{12} = \frac{1}{q_{l-1}}$$

$$\implies a = \sqrt{\frac{3}{q_{l-1}}}$$

Backward pass



$h^l \in R^{q_l}$

$$\frac{\partial \hat{y}}{\partial h_i^l} = \sum_{j=0}^{q_{l+1}-1} \frac{\partial \hat{y}}{\partial h_j^{l+1}} \underbrace{\frac{\partial h_j^{l+1}}{\partial h_i^l}}_{w_{ji}^l}$$

$$V\left[\frac{\partial \hat{y}}{\partial h_i^l}\right] = V\left[\sum_{j=0}^{q_{l+1}-1} \frac{\partial \hat{y}}{\partial h_j^{l+1}} w_{ji}^{l+1}\right]$$

$$= \sum_{j=0}^{q_{l+1}-1} V\left[\frac{\partial \hat{y}}{\partial h_j^{l+1}}\right] V\left[w_{ji}^{l+1}\right]$$

$$V\left[\frac{\partial \hat{y}}{\partial h^{l}}\right] = q_{l+1}\, V\left[\frac{\partial \hat{y}}{\partial h^{l+1}}\right] V\left[w_{ji}^{l+1}\right]$$

$$q_{l+1}\, V\left[w_{ji}^{l+1}\right] = 1 \implies V\left[w_{ji}^{l+1}\right] = \frac{1}{q_{l+1}}$$

$$V\left[w^{l}\right] = \frac{1}{q_{l-1}}$$

$$V\left[w^{l}\right] = \frac{1}{q_{l}}$$

Xavier initialization:

$$V\left[w^{l}\right] = \frac{1}{\frac{q_{l-1}+q_{l}}{2}}$$

$$w_{ij}^{l} \sim U\left[-\sqrt{\frac{6}{q_{l-1}+q_{l}}},\ \sqrt{\frac{6}{q_{l-1}+q_{l}}}\right]$$