Роль начального приближения.

Дисперсия карт активаций

SGD

1. $\Theta_{t=0}$, η, B

2. $\Theta_{t=t+1} = \Theta_t - \eta \nabla_\Theta \mathcal{L}(B)$

3. Stop $\Rightarrow \Theta^*$

$z^{(1)} = \Theta^{T(1)} X$

$h^{(1)} = g(z^{(1)})$

$z^{(2)} = \Theta^{T(2)} h^{(1)}$

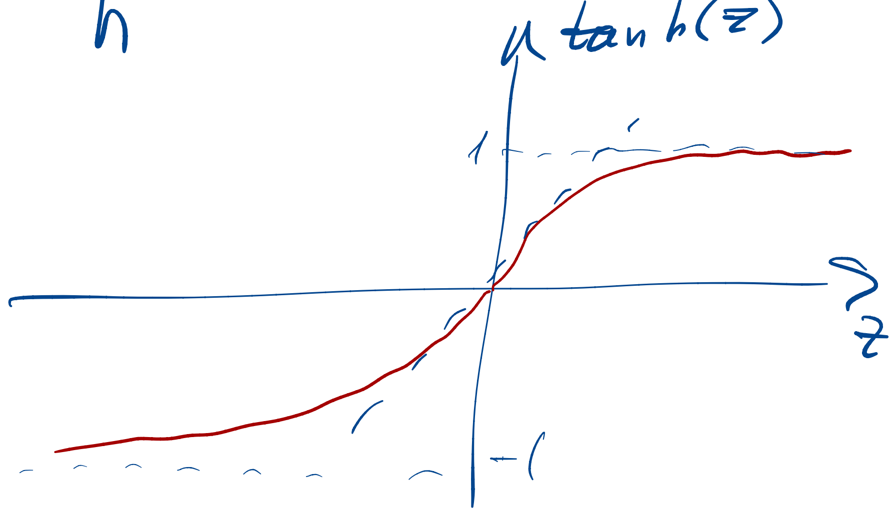
$h^{(2)} = g(z^{(2)})$

$z^{(3)} = \Theta^{T(3)} h^{(2)}$

$y = \psi(z^{(3)})$

$$x \to \underbrace{\boxed{\Theta^T x}}_{z^{(1)}} - \underbrace{\boxed{g(z)}}_{h^{(1)}} - \underbrace{\boxed{\Theta^T h^{(1)}}}_{z^{(2)}} - \underbrace{\boxed{g(z^{(2)})}}_{h^{(2)}} \cdots$$

$$\cdots - \underbrace{\boxed{\Theta^{T(\ell)} h^{(\ell-1)}}}_{z^{(\ell)}} - \underbrace{\boxed{g(z^{(\ell)})}}_{h^{(\ell)}} - \cdots - \hat{y}$$



$g(z) = \tanh(z)$

$z \approx 0 \Rightarrow \tanh(z) \approx z$

① $x_i \sim$ i.i.d.

$h_i \sim$ i.i.d.

② $\Theta_i^{(\ell)} \sim$ i.i.d.

$\Theta^{(\ell)} : \; n \times m$

③ $x \perp \Theta$

$h \perp \Theta$

$$\mathrm{Var}(z_i) = \mathrm{Var}(\theta_i h_i) =$$

$$= \mathbb{E}(\theta_i h_i)^2 - (\mathbb{E}(\theta_i h_i))^2 =$$

$$= \mathbb{E}\,\theta_i^2 h_i^2 - (\mathbb{E}\,\theta_i h_i)^2 =$$

$$(\mathbb{E}\,\theta_i h_i)^2 = (\mathbb{E}\,\theta_i\, \mathbb{E}\, h_i)^2$$

$$= \mathbb{E}\,\theta_i^2\, \mathbb{E}\, h_i^2 - (\mathbb{E}\,\theta_i)^2 (\mathbb{E}\, h_i)^2 =$$

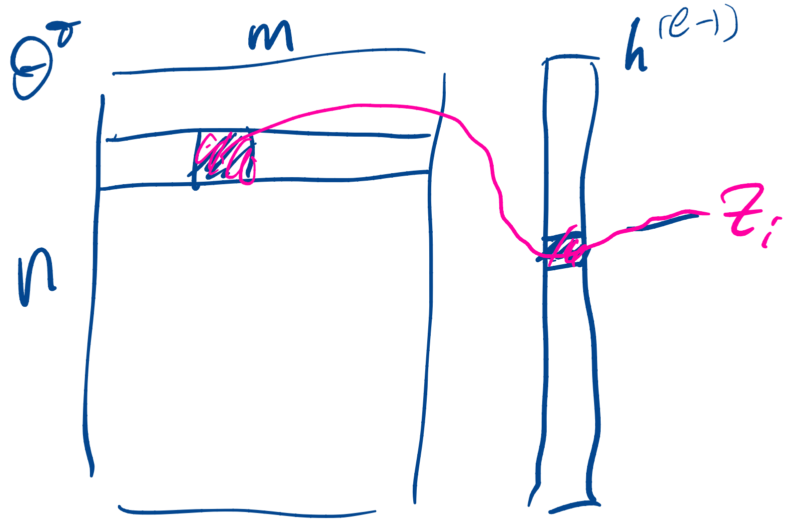
$$= \left(\mathbb{E}\,\theta_i^2 - (\mathbb{E}\,\theta_i)^2 + (\mathbb{E}\,\theta_i)^2\right)\left(\mathbb{E}\, h_i^2 - (\mathbb{E}\, h_i)^2 + (\mathbb{E}\, h_i)^2\right)$$

$$- (\mathbb{E}\,\theta_i)^2 (\mathbb{E}\, h_i)^2 =$$

$$= \left(\mathrm{Var}\,\theta_i + \cancelto{0}{(\mathbb{E}\,\theta_i)^2}\right)\left(\mathrm{Var}\, h_i + \cancelto{0}{(\mathbb{E}\, h_i)^2}\right) -$$

$$- \cancelto{0}{(\mathbb{E}\,\theta_i)^2}\, \cancelto{0}{(\mathbb{E}\, h_i)^2} =$$

$$= \mathrm{Var}\,\theta_i\ \mathrm{Var}\, h_i$$



$$z_i = \theta_i^{(\ell)} h_i$$

$$z = \sum_{i=1}^{m} \theta_i h_i$$

$$\mathrm{Var}\, z = \sum_{i=1}^{m} \mathrm{Var}\,\theta_i \,\mathrm{Var}\, h_i = m \,\mathrm{Var}\,\theta \,\mathrm{Var}\, h$$

$$g(z) \approx z$$

$$\mathrm{Var}(g(z)) \approx \mathrm{Var}\, z = m \,\mathrm{Var}\,\theta\,\mathrm{Var}\, h$$

$$\mathrm{Var}\, h^{(\ell+1)} \approx \mathrm{Var}\, h^{(\ell)}$$

$$\begin{array}{ccc} \| & & \| \\ g(z^{(\ell+1)}) & & g(z^{(\ell)}) \end{array}$$

$$m \,\mathrm{Var}\,\theta\, \cancel{\mathrm{Var}\, h^{(\ell)}} \approx \cancel{\mathrm{Var}\, h^{(\ell)}}$$

$$\mathrm{Var}\,\theta \approx \frac{1}{m}$$

$\theta \sim U[a; b]$ $\quad a = -b$

$$\mathbb{E}\,\theta = \int_a^b \underbrace{p(\theta)}_{c}\,\theta\, d\theta = \int_a^b c\,\theta\, d\theta = c\,\frac{\theta^2}{2}\Big|_a^b =$$

$$= c\left(\frac{b^2}{2} - \frac{a^2}{2}\right)$$

$$\int_a^b p(\theta)\, d\theta = 1 = \int_a^b c\, d\theta = c\,(b-a) = 1$$

$$c = \frac{1}{b-a} = \frac{1}{2a}$$

T.B.C.