Xavier Glorot

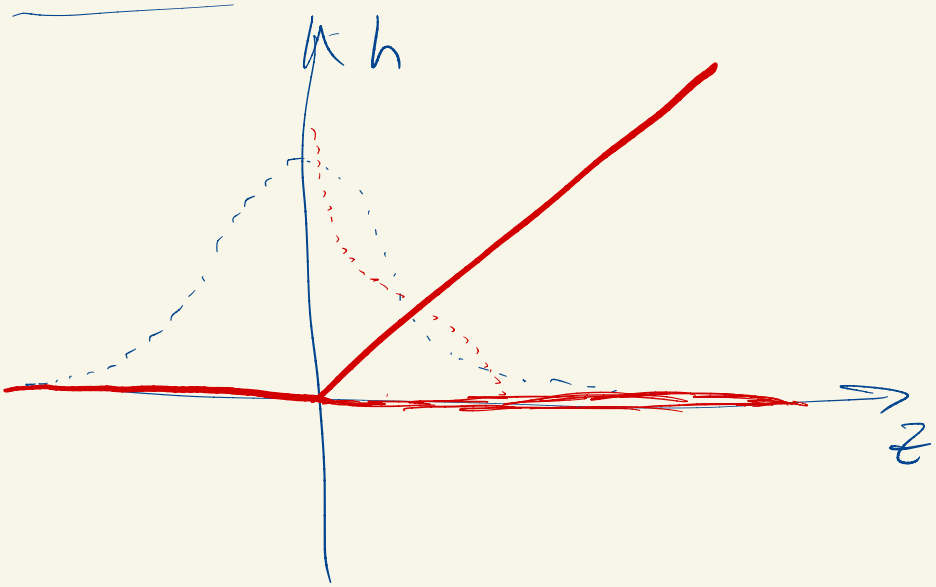
$$\theta \sim \mathcal{N}(\mu=0, \sigma^2) \qquad \sigma^2(\theta) = \frac{1}{n_{in}+k+1}$$

$$\mathcal{U}(-a, a)$$

$$\sigma^2(a) = \int_{-a}^{a} (\theta-\mu)^2 p(\theta)\, d\theta \qquad a = \dots \; (\sigma)$$

Kaiming He

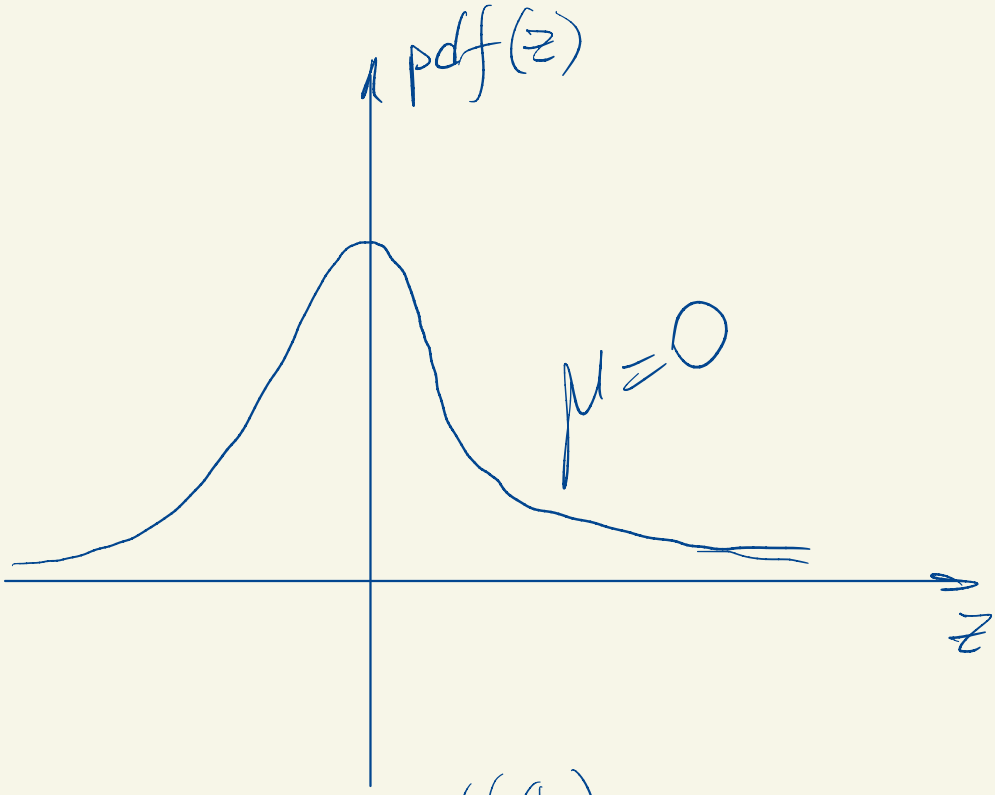
$$\sigma^2 = \frac{2}{k}$$

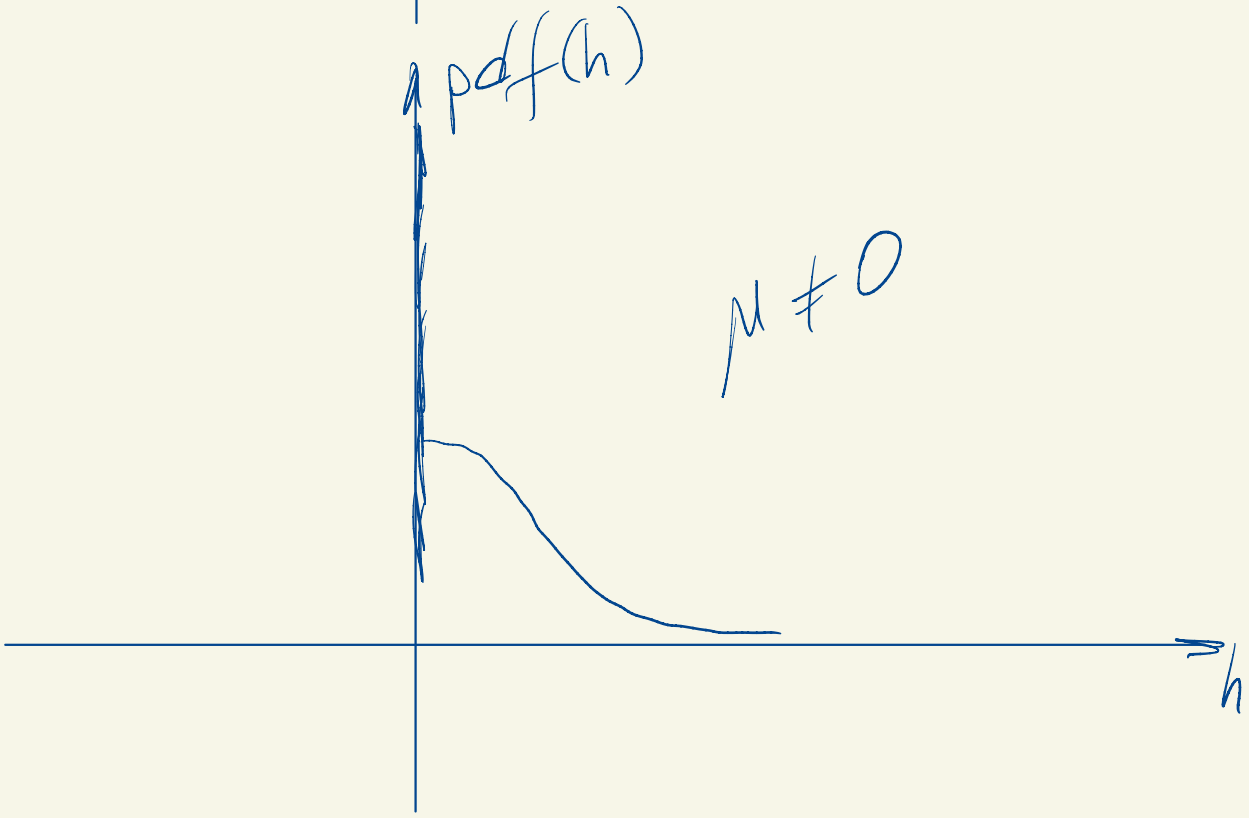


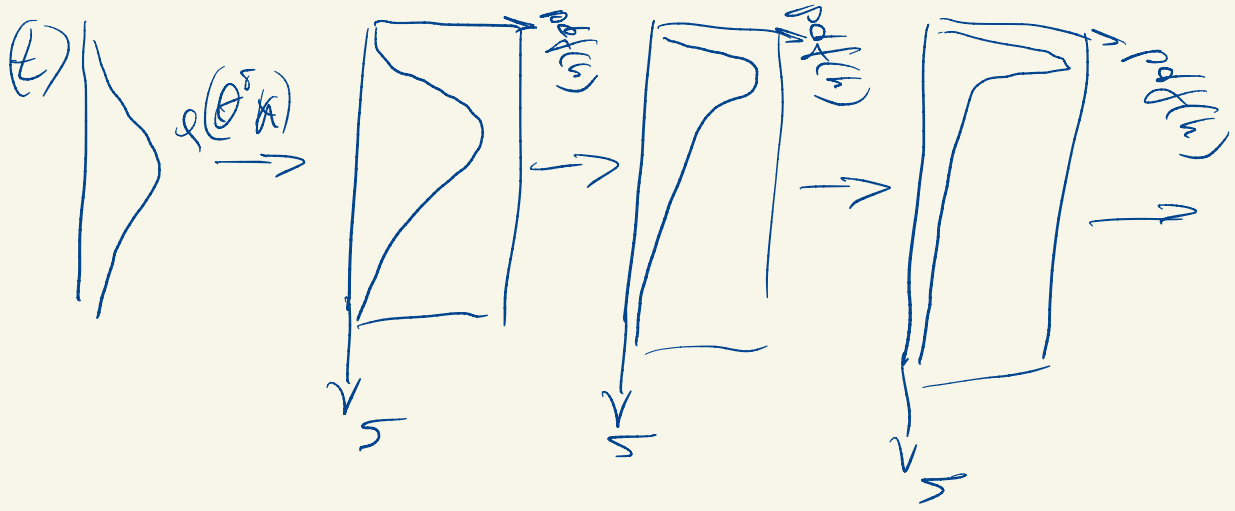
$$z^{(e)} = \theta^T h^{(e-1)}$$

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

$h = ReLU(z)$







(t+1)

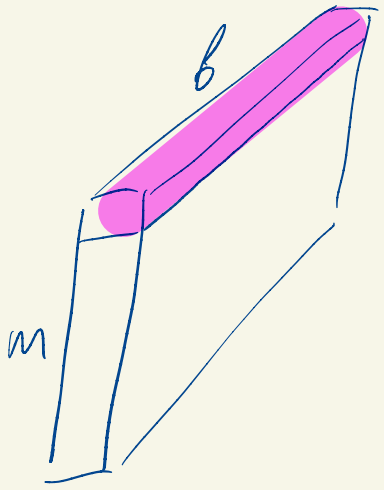
$$\ell^* = \ell + \lambda \psi(\theta)$$

$\mu(\theta)$

$\sigma^2(\theta)$

$$z^{(l)} = \theta^T h^{(l-1)}$$

$$z^* = \frac{z - \mu_B(z)}{\sigma_B(z)}$$



$$\sigma(z^*) = 1$$

$$\mu(z^*) = 0$$

$$\mu_t^* = \mu_{t-1} \beta + \mathbb{E}_{B_t}(z)$$

$$\sigma_t^* = \sigma_{t-1} \alpha + \sigma_{B_t}(z)$$

$$z^* = \frac{z - \mu_t^*(z)}{\sigma_t^*(z) + \varepsilon} \cdot \gamma + \xi$$

$\gamma = 1$, $\xi = 0$

Пакетная нормализация

Batch Normalization

Batch Norm