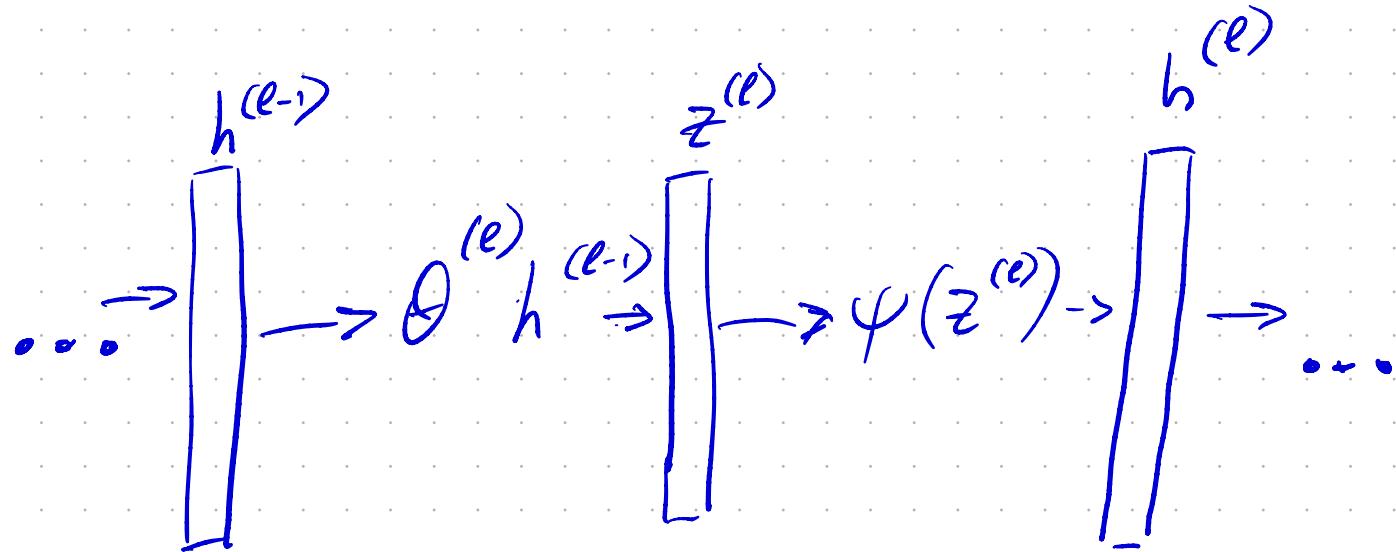


Нормализация (Batch Normalization)

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



$$z^* = \frac{z - \mu^{(e)}}{\sigma^{(e)} + \epsilon}$$

$$\mu = E(z)$$

$$\sigma = \sqrt{Var(z)}$$

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

"momentum"

$$\beta = 0,1 \text{ by default}$$

$$\sigma_t^2 = (1 - \beta) \sigma_{t-1}^2 + \beta \sigma_{t-1}^2$$

$$z^* = \frac{z^{(e)} - \mu^{*(e)}}{\sqrt{\sigma^2} + \epsilon} \cdot \gamma + \beta$$

Batch Norm: 3 art für den normale

$z: d$

$z \in \mathbb{R}^d$

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

$\mu^{*(l)}: d$

$$G_t^{z*(l)} = (1 - \beta) G_{t-1}^{z*(l)} + \beta z$$

$G^{z*(l)}: d$

$$z^{*(l)} = \frac{z^{(l)} - \mu^{*(l)}}{\sqrt{G^{z*(l)}} + \epsilon}$$

$z^*: d$

Batch Normalization

Pros

BN может ускорять обучение за счет стандартизации

Но BN можно не сильно контролировать производство



BN эффективно улучшает производство функции потерь

BN работает как регуляризатор

Cons

BN плохо работает при batch_size недостаточных ($n \sim 1-5-10$)

BN побуждает к перебору параметров

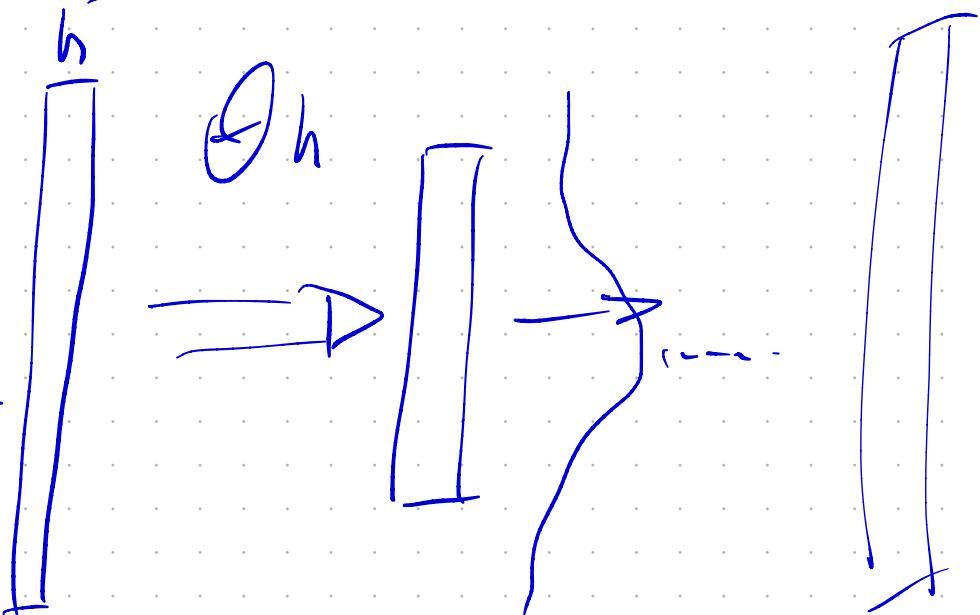
BN неизбежно использует

в загораживающих коммуникациях

поставляемое устройством

из разных производителей

Dropout Преобразование



$$P.\text{shape} = \Theta.\text{shape}$$

$$P \sim \text{Ber}(p)$$

$$\Theta^* = \Theta \odot P \quad z^{(l)} = \Theta^{*(l)} h^{(l-1)}$$

$$E(z^*) = E(\Theta^* h) = E(\Theta^*) E(h)$$

$$\Theta^* = \Theta \odot P \quad E(\Theta^*) = E(\Theta) E(P) = p E(\Theta)$$

$$E(z^*) = p E(\Theta) E(h) = p E(z)$$

Dropout

Люб

но этаке Enference

$$z^{**} = z^* - \rho$$

Люб

на этаке обуления

$$z^{**} = z^*/\rho$$