



deeplearning.ai

# Batch Normalization

---

## Batch Norm at test time

(During test, we may not have minibatches, Batch norm finds mean & variance per minibatch,  $\Rightarrow$  If we have a single test sample, how do we use Batch norm?)

# Batch Norm at test time

$M = \# \text{ examples / minibatch}$

$$\mu = \frac{1}{m} \sum_i z^{(i)}$$

$$\sigma^2 = \frac{1}{m} \sum_i (z^{(i)} - \mu)^2$$

Training time

$$z_{\text{norm}}^{(i)} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2 + \epsilon}}$$

$$\tilde{z}^{(i)} = \gamma z_{\text{norm}}^{(i)} + \beta$$

- $\tilde{z}^{(i)}$  and  $z_{\text{norm}}^{(i)}$  are calculated in training time & depend on  $\mu$  &  $\sigma^2$  which depend on minibatch size "m"
  - In test, you may not have m samples. You have 1 sample to give the o/p to
  - But we still need  $\mu$  &  $\sigma^2$  in test  $\Rightarrow$  we estimate it using exponentially weighted args (across all minibatches)
- pick some layer "l"
- |  |   |
|--|---|
| $x^{\langle 1 \rangle}, x^{\langle 2 \rangle}, x^{\langle 3 \rangle}, \dots$                   | $x^{\langle 1 \rangle}, x^{\langle 2 \rangle}, x^{\langle 3 \rangle}, \dots$                                  |
| $\downarrow \quad \downarrow \quad \downarrow$   | $\downarrow \quad \downarrow \quad \downarrow$  |
| $\mu^{\langle 1 \rangle [l]}, \mu^{\langle 2 \rangle [l]}, \mu^{\langle 3 \rangle [l]}, \dots$ | $\sigma^2^{\langle 1 \rangle [l]}, \sigma^2^{\langle 2 \rangle [l]}, \sigma^2^{\langle 3 \rangle [l]}, \dots$ |
- $\downarrow$  mean of 1<sup>st</sup> minibatch on layer l
- To find  $\mu^{\langle 1 \rangle [l]}, \dots$  &  $\sigma^2^{\langle 1 \rangle [l]}, \dots$   
we maintain moving averages or exponentially weighted args
- Say we call moving Arg =  $\mu', \sigma'^2$
- then in test time  $z_{\text{norm}} = \frac{z - \mu'}{\sigma'^2}$