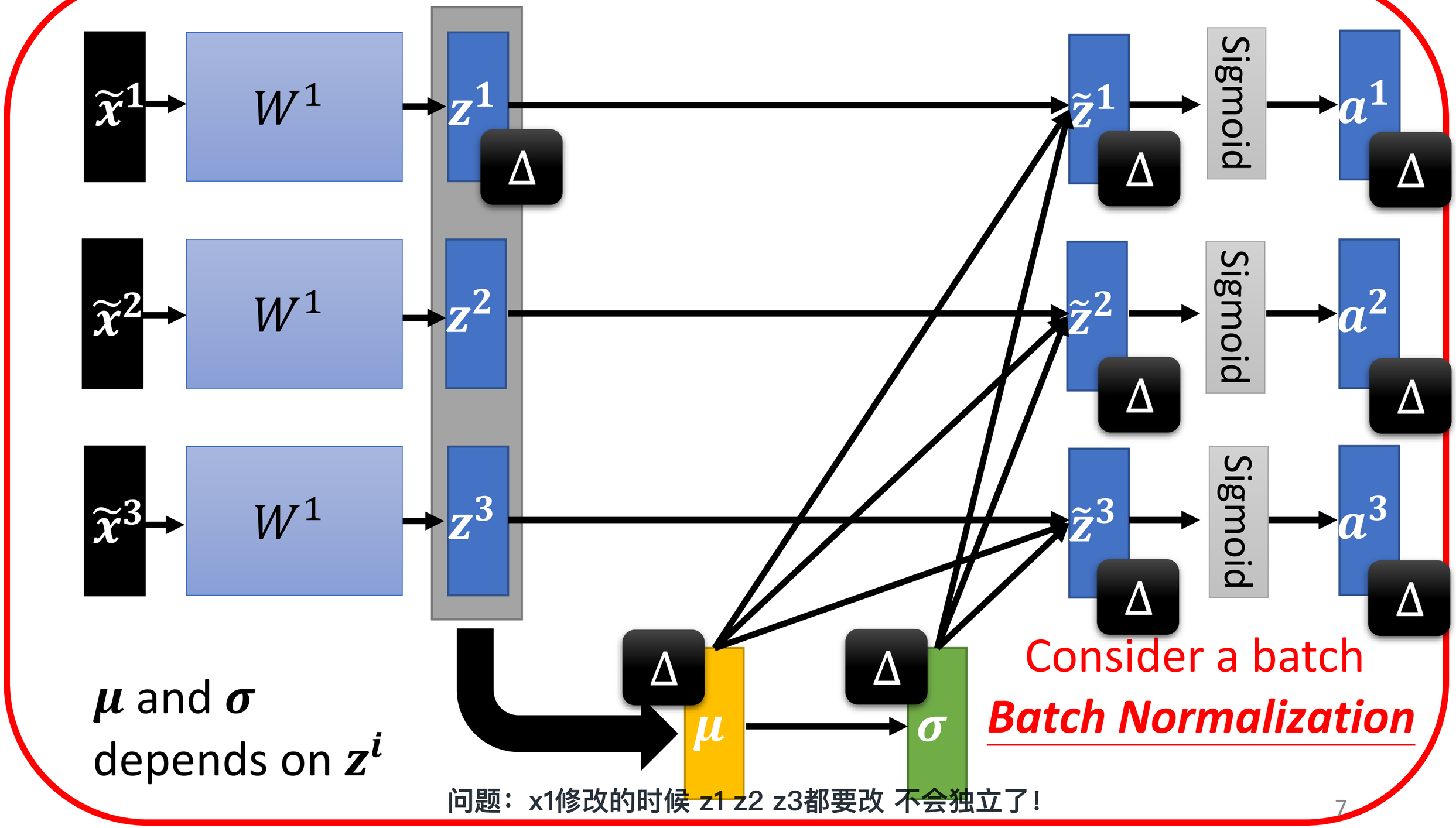
Considering Deep Learning

$$\tilde{z}^i = \frac{z^i - \mu}{\sigma}$$

This is a large network!



μ and σ depends on z^i

Consider a batch

Batch Normalization

问题：x1修改的时候 z1 z2 z3都要改 不会独立了！

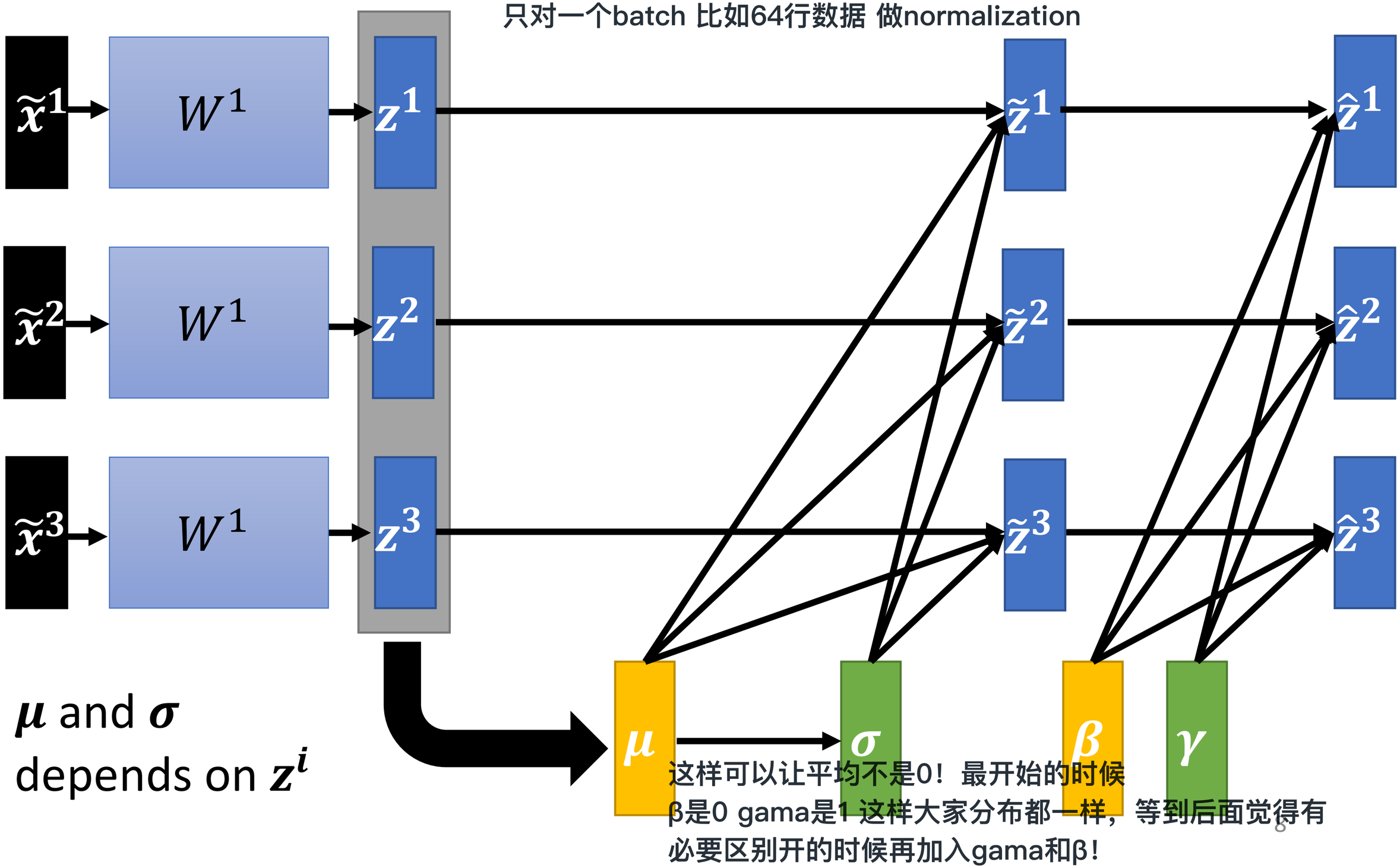
Batch normalization

$$\tilde{z}^i = \frac{z^i - \mu}{\sigma}$$

element wise两两相乘

$$\hat{z}^i = \gamma \odot \tilde{z}^i + \beta$$

只对一个batch 比如64行数据 做normalization

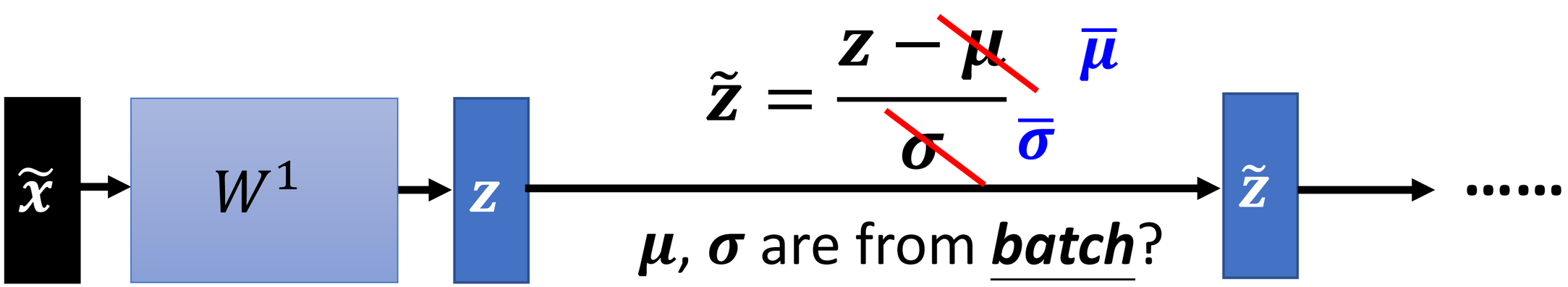


μ and σ depends on z^i

这样可以让平均不是0！最开始的时候β是0 gama是1 这样大家分布都一样，等到后面觉得有必要区别开的时候再加入gama和β！

Batch normalization – Testing

线上的时候！每一行data进来就要做一次运算



testing的时候！

We do not always have ***batch*** at testing stage.

Computing the moving average of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ of the batches during training.

$$\boldsymbol{\mu}^1 \quad \boldsymbol{\mu}^2 \quad \boldsymbol{\mu}^3 \quad \ldots\ldots \quad \boldsymbol{\mu}^t$$

$$\bar{\boldsymbol{\mu}} \leftarrow p\bar{\boldsymbol{\mu}} + (1-p)\boldsymbol{\mu}^t$$

moving average，p默认是0.1