

Batchnorm

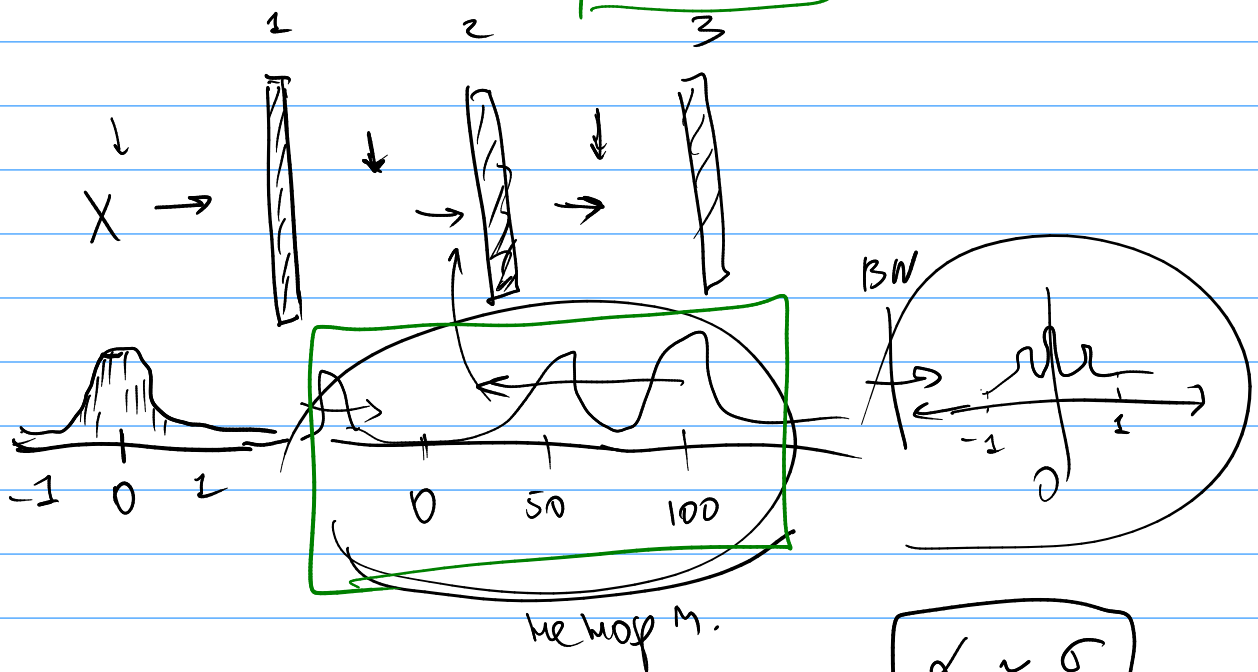
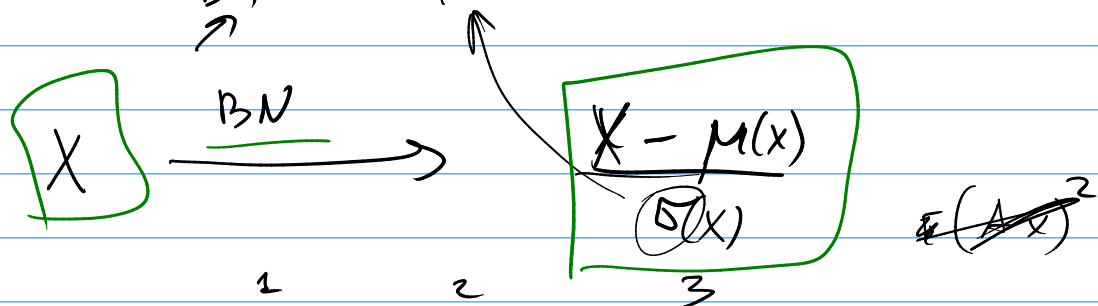
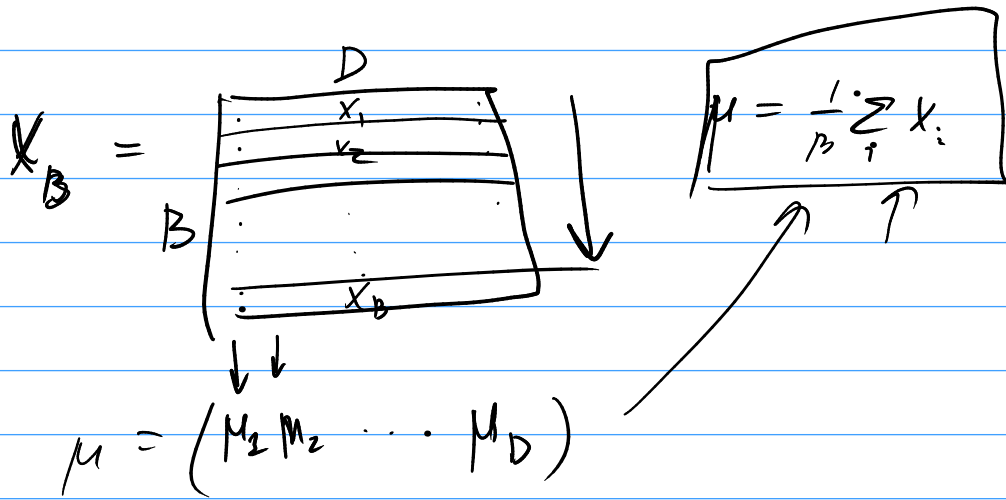


Diagram illustrating the Batch Normalization (BN) operation with parameters α and β .

Input: X

Parameters: $\alpha \sim \sigma$, $\beta \sim \mu$

BN operation: $BN(\alpha, \beta) = \frac{\alpha}{\sigma} \frac{X - \mu(x)}{\sigma} + \beta$

Result: X

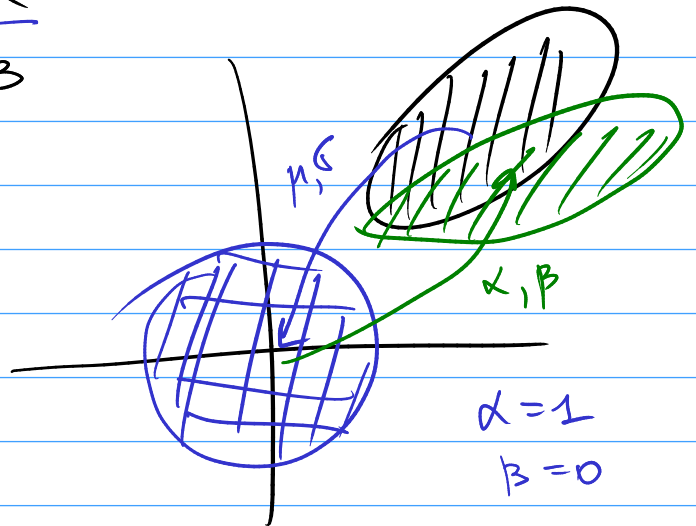
$$\underline{\beta = 1}$$

$$\begin{array}{l} \sigma \rightarrow \alpha \\ \mu \rightarrow \beta \end{array}$$

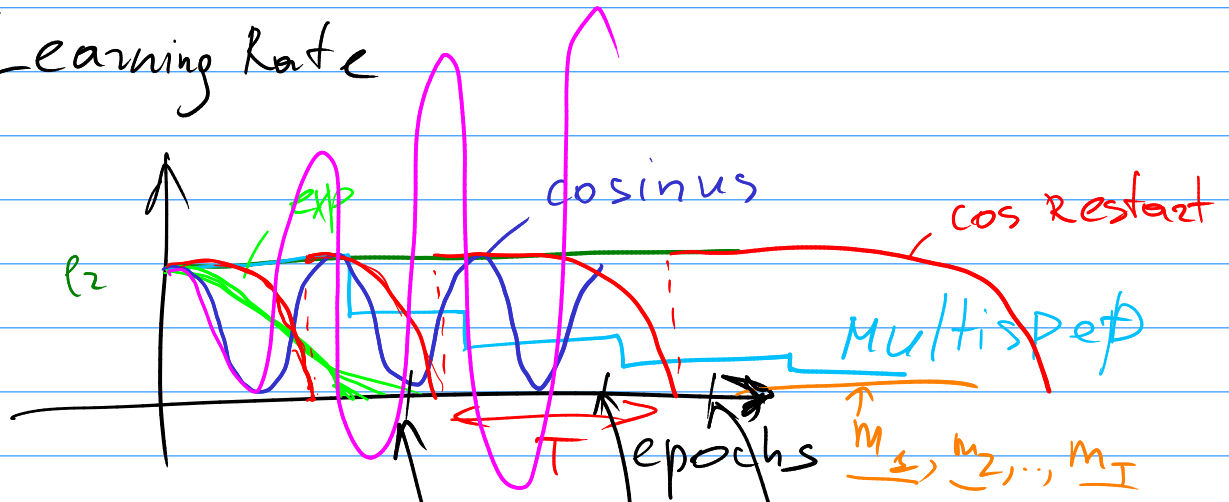
$$\underline{Y = \alpha X + \beta}$$

train

eval



Learning Rate



$$l_2 = \cos(\text{epoch})$$

$$\begin{array}{l} m_i(x) = y_i \\ y = \frac{1}{T} \sum_{i=1}^T y_i \end{array}$$

ensembling

Averaged SGD

$$m \equiv \frac{1}{K} \sum_{i=1}^K m_i$$

$$\underline{m(x) = y}$$

Auto encoder

