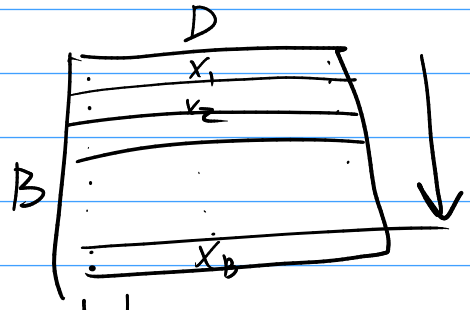
Batchnorm

$X_B =$



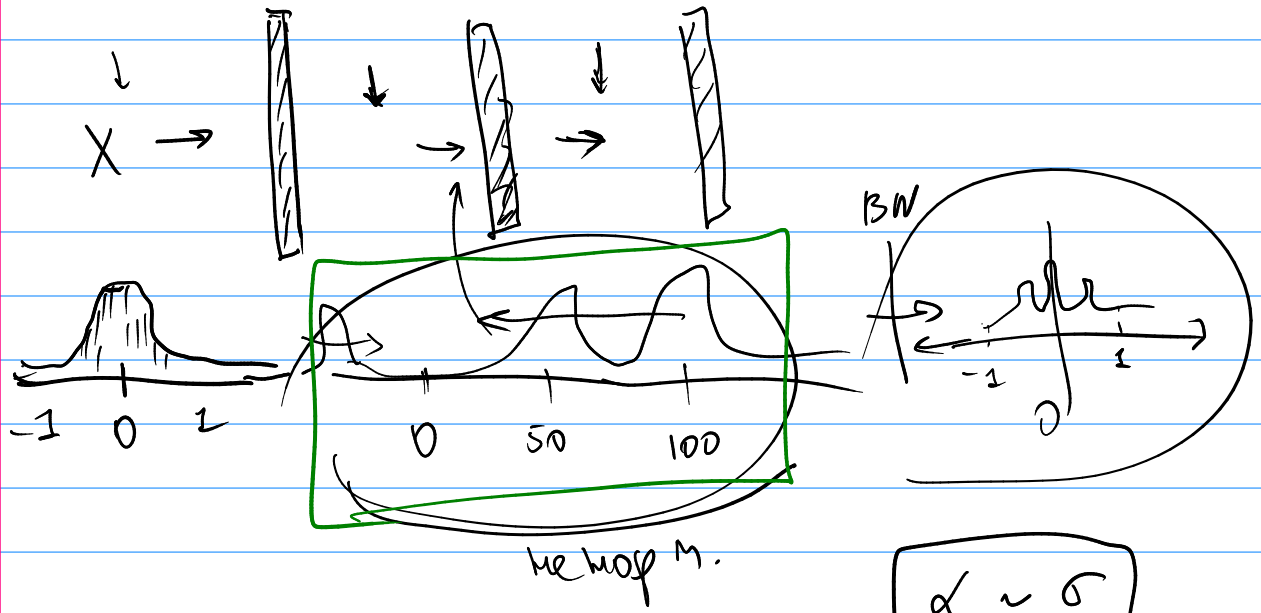
$$\boxed{\mu = \frac{1}{B}\sum_i x_i}$$

$$\mu = (\mu_1\ \mu_2\ \cdots\ \mu_D)$$

$$\sigma = \frac{1}{B-1}\sum_{i=1}^{B}(x_i - \mu)^2 = (\sigma_1, \ldots, \sigma_D)$$

$$\boxed{X} \xrightarrow{BN} \boxed{\frac{X - \mu(x)}{\sigma(x)}}$$

$\cancel{\sigma(Ax)^2}$



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

$\boxed{B = 1}$

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

$$= \cancel{\sigma}\,\frac{x - \cancel{\mu}}{\cancel{\sigma}} + \cancel{\mu} = X$$

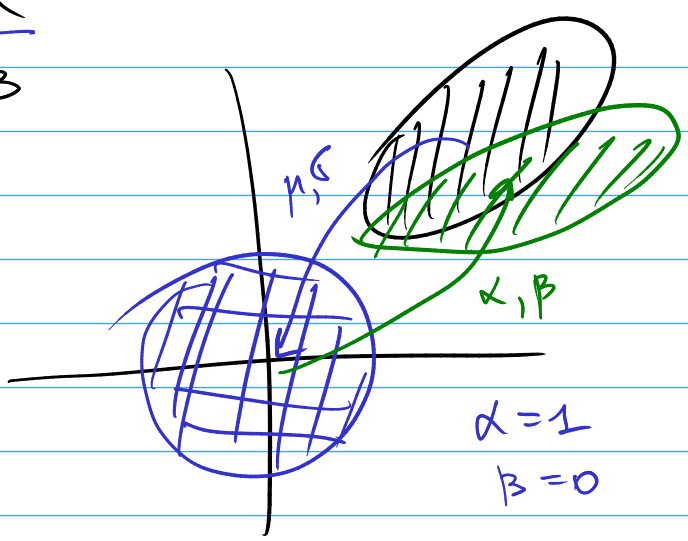
$\beta = 1$ $\quad \sigma \to \alpha$

$\mu \to \beta$

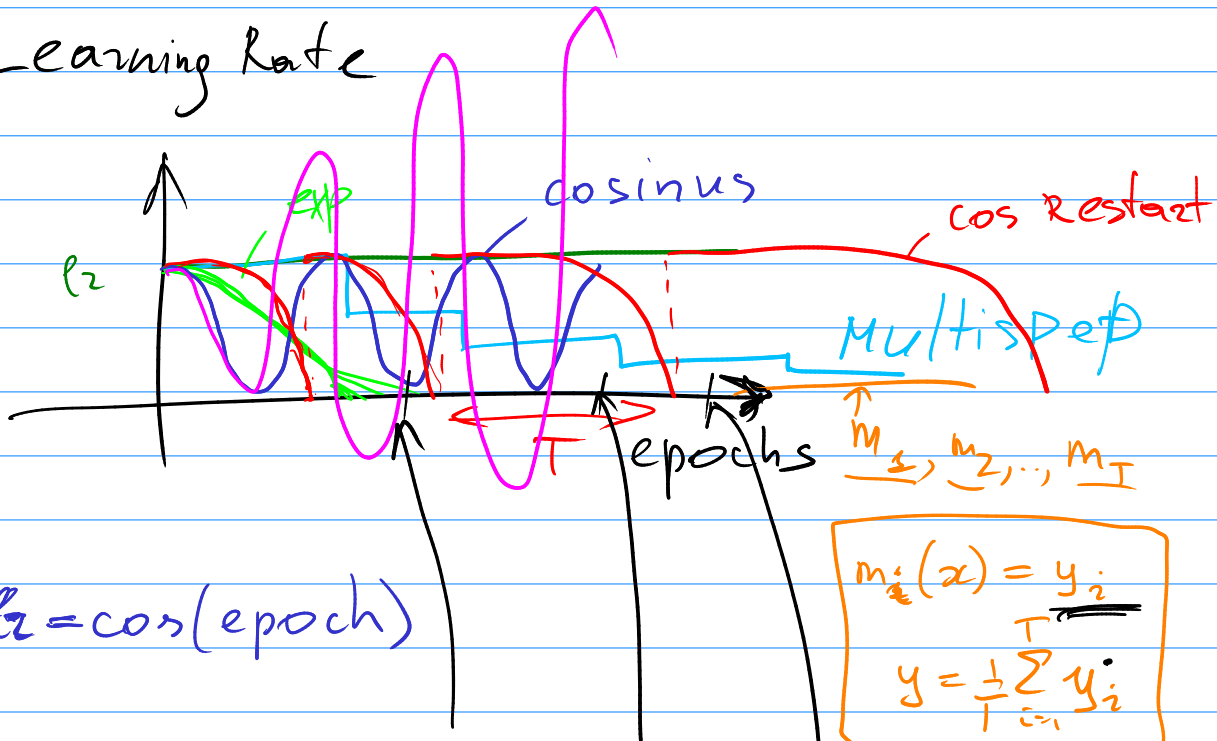
$Y = \alpha X + \beta$

t_{train}

eval



Learning Rate



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

$m_i(x) = y_i$

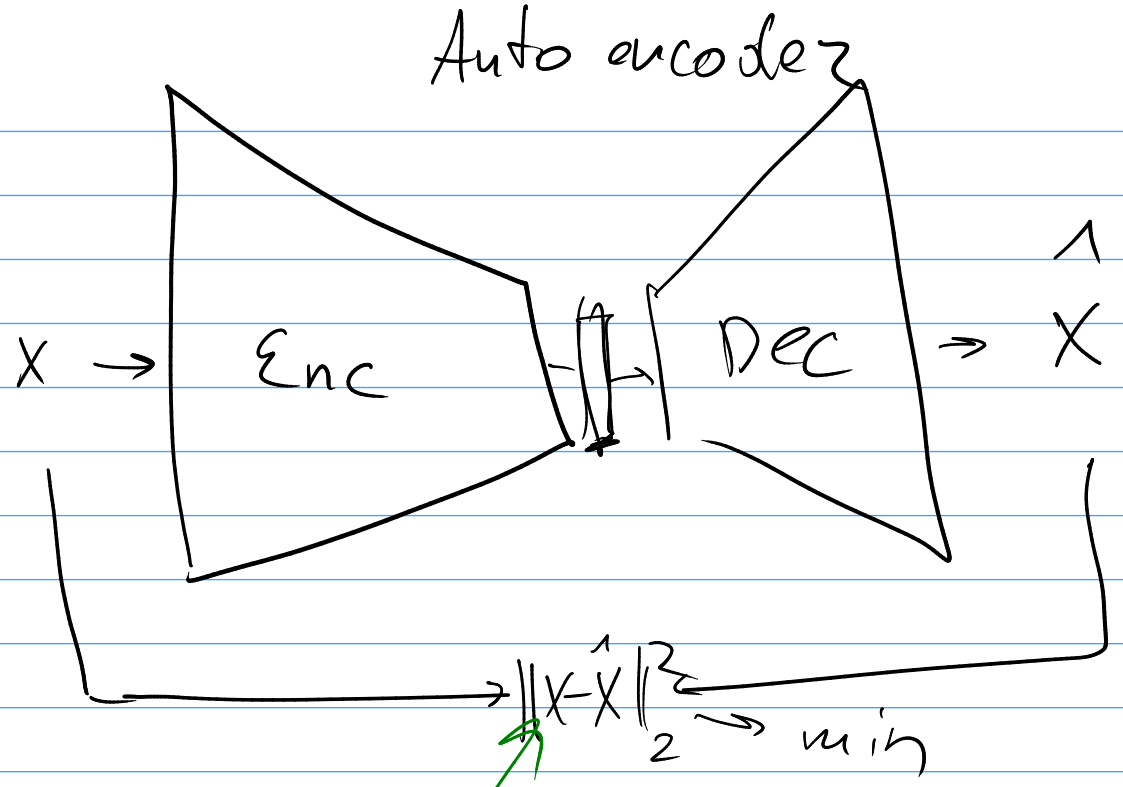
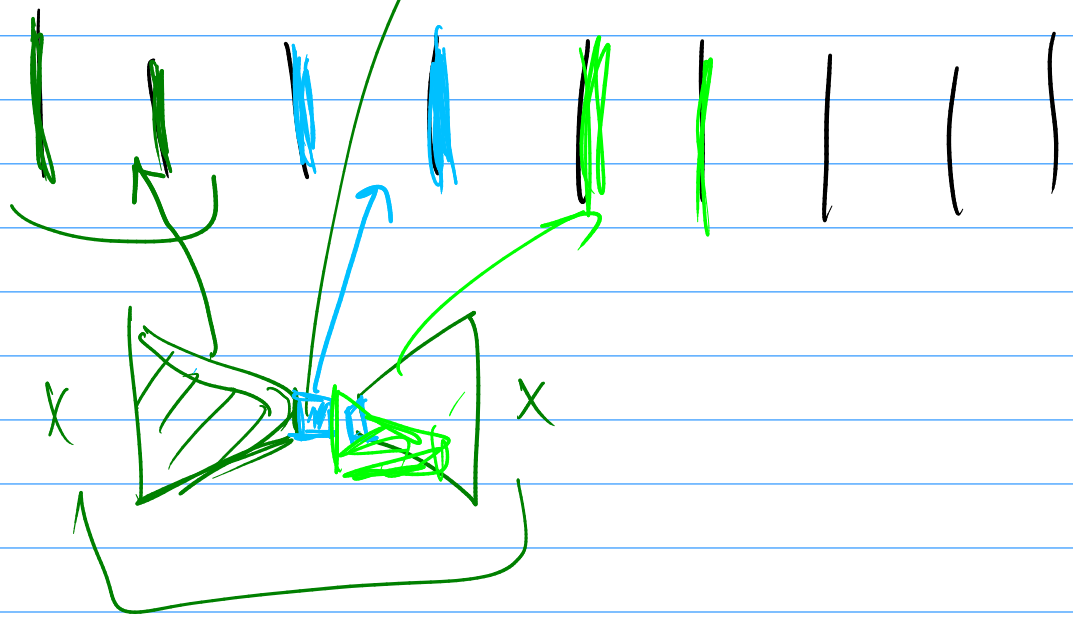
$y = \frac{1}{T} \sum_{i=1}^{T} y_i$

ensembling

m_1 $\quad m_2$ $\quad m_K$

Averaged SGD

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

$m(x) = y$

Auto encoder