2025-08-04 - Compare and Contrast among activation functions, loss functions, optimizers and regularization for choosing the appropriate method for the given application

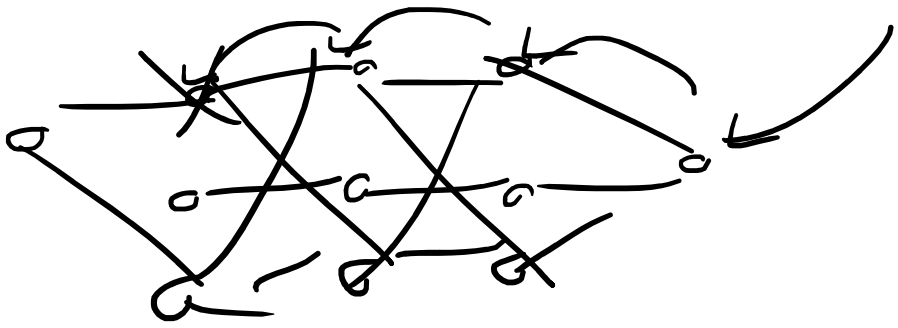
05 August 2025 11:34

=> Overfitting Reducing Techniques

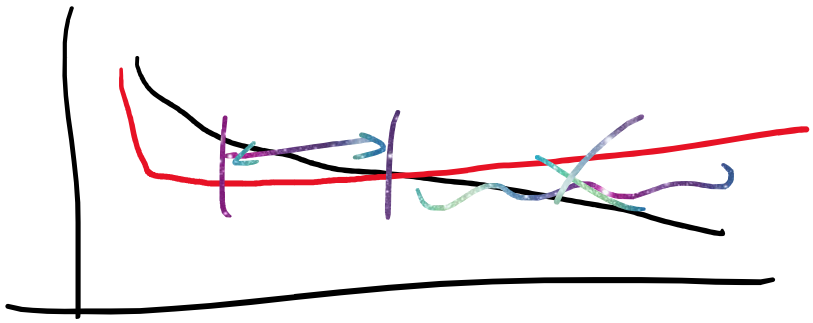
-> Regularization
-> Early Stopping
-> Dropout
-> Batch Normalization

=> Early Stopping : Breaks the cycle of model being overfit.

Vanishing Gradients $\partial/\partial w$ $\partial/\partial b$

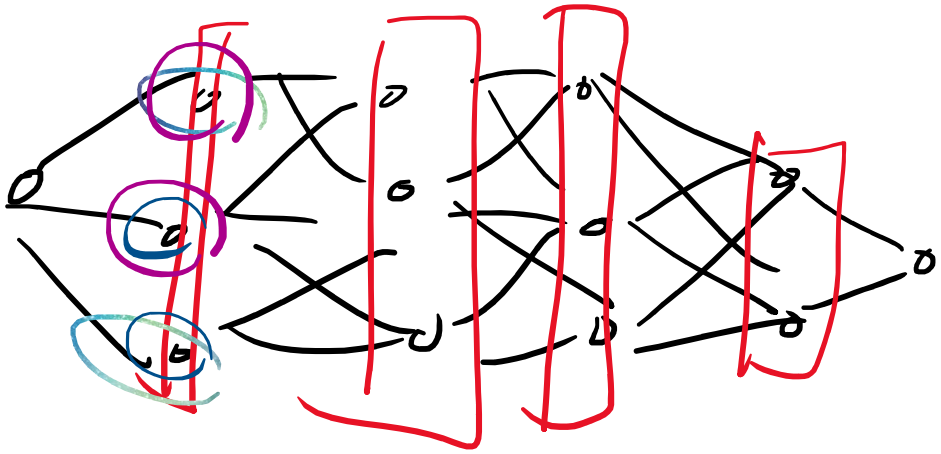


Back-prop



=> Dropout :

activations



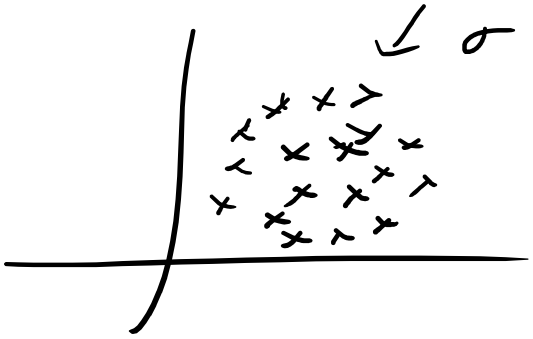
LI/. ✓

25-50

=> Batch Normalization :

$\mu = n$

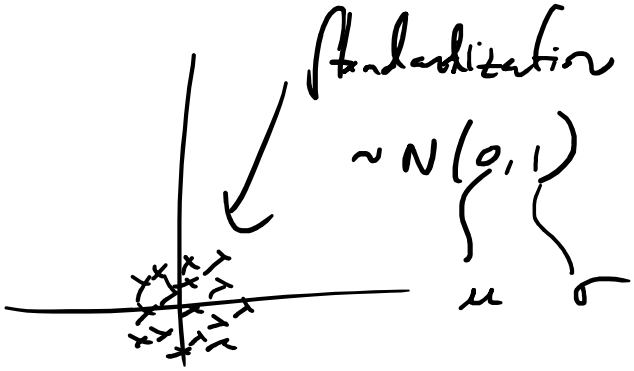
$\sigma = n$

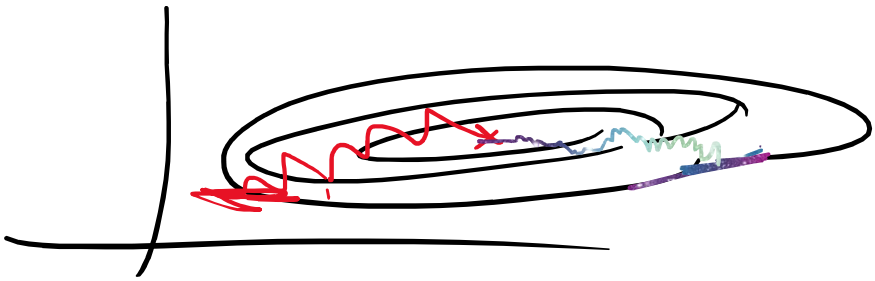


standardization

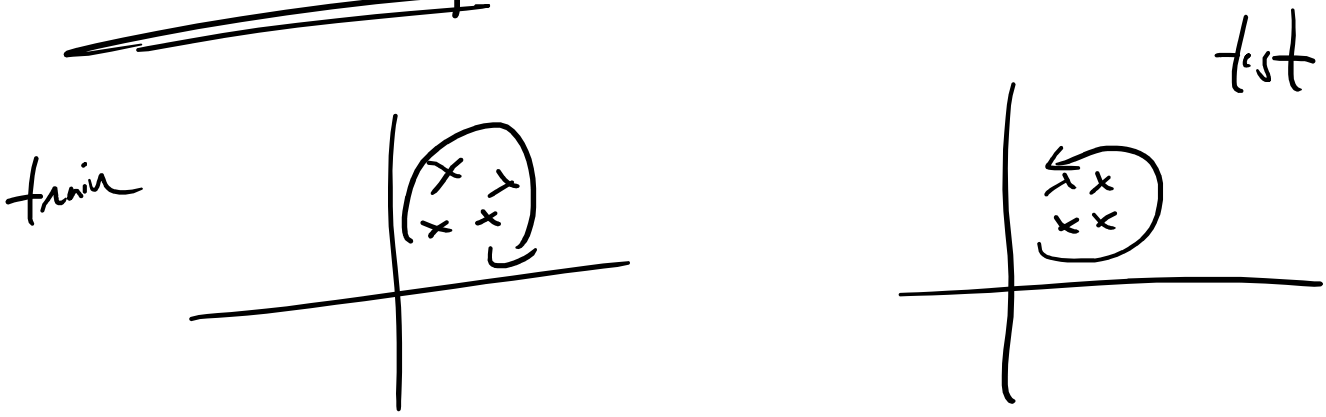
$\sim N(0,1)$

μ σ





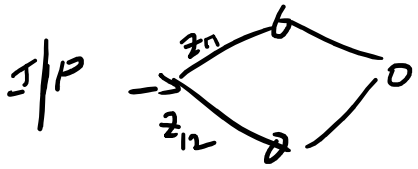
=> Covariate Shift :



$a = g(z)$ ↳ Predict Fn

Activation layer ↓ Activation fn (Mathematical Fn)

$(z = w \cdot x + b)$ Normalize/Standardize z First



$z_{11} \rightarrow z_{11}^{N} \rightarrow g(z_{11}^{N}) = a_{11}$ ✓

$$\Rightarrow \boxed{z = \frac{x - a}{\sigma}} \approx \frac{z_{11} - \mu}{\sigma} \quad (\mu \leftarrow 0,\ \sigma \leftarrow 1)$$

$\Rightarrow \quad z_{11} \rightarrow z_{11}^{N} \rightarrow z_{11}^{BN} \rightarrow g(z_{11}^{BN}) = a_{11}$

Batches are used to utilize memory effectively.