

2025-08-05 - Compare and Contrast among activation functions, loss functions, optimizers and regularization for choosing the appropriate method for the given application.

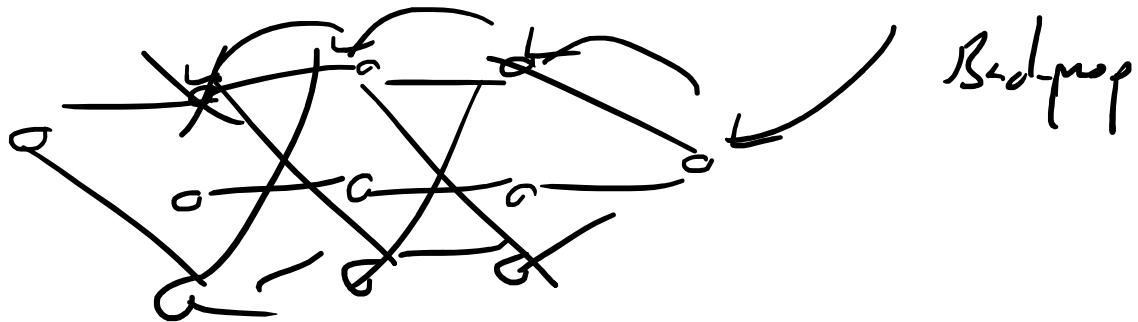
05 August 2025 11:37

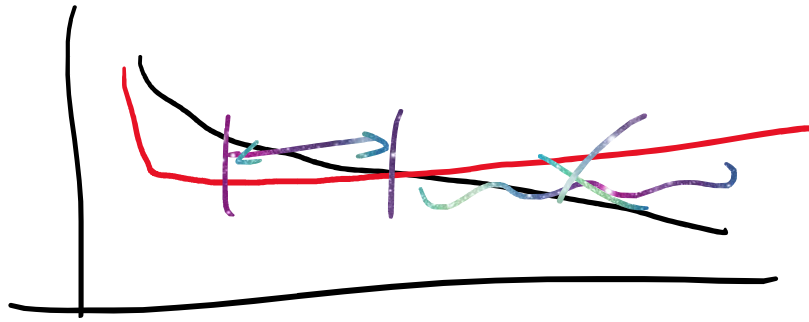
Overfitting Reducing Techniques

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

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

Vanishing Gradients d/dw d/db

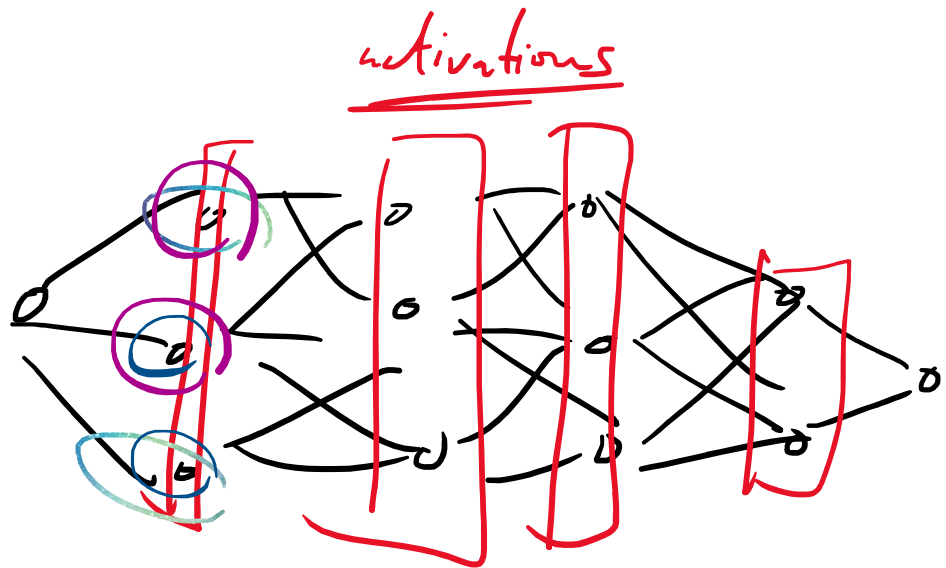




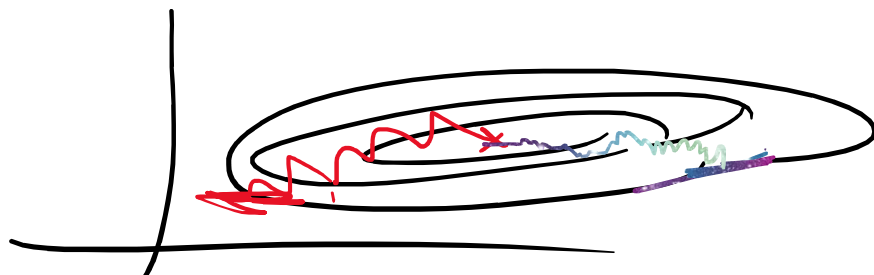
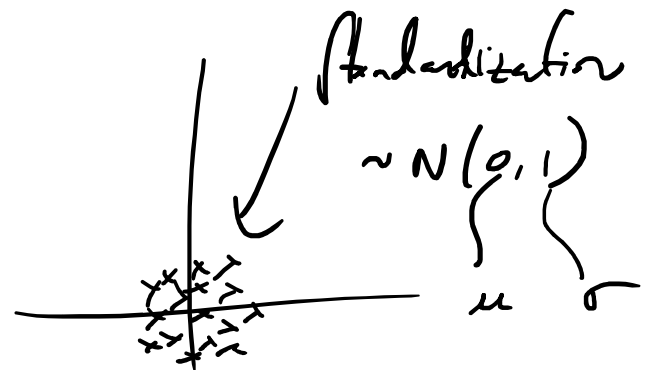
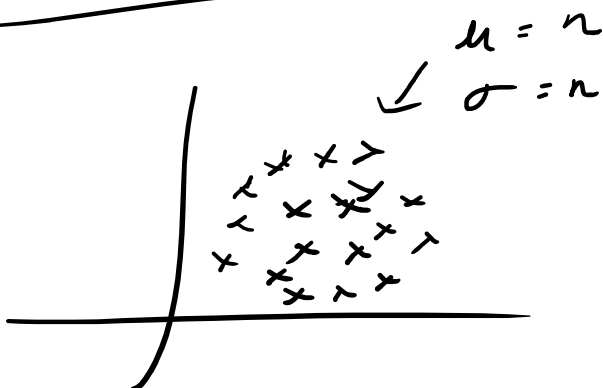
\Rightarrow Dropout :

25% ✓

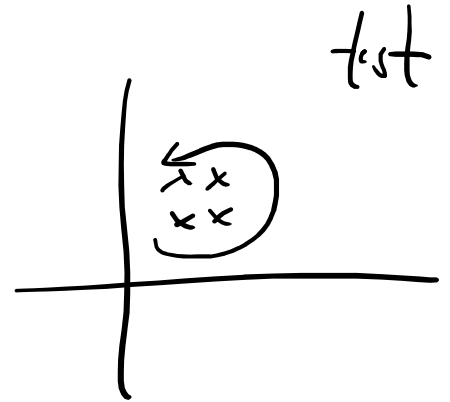
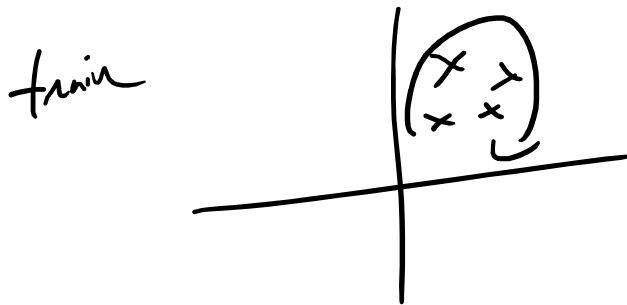
25-50



\Rightarrow Batch Normalization :

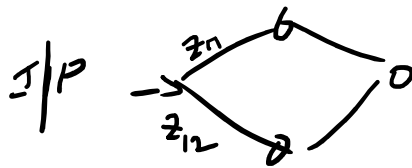


\Rightarrow Covariate Shift :



$\hat{a} = g(z)$ Predict Fn
 Activation layer \downarrow Activation Fn (Mathematical Fn)

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



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

$$\Rightarrow z = \frac{\bar{z} - \mu}{\sigma}$$

$$\approx \frac{z_{11} - \mu}{\sigma}$$

$\nwarrow \nearrow$
0 1

$$\Rightarrow z_{ii} \rightarrow z_{ii}^N \rightarrow z_{ii}^{BN} \rightarrow g(z_{ii}^{BN}) = \underline{a_{ii}}$$

Batches are used to utilize memory effectively.