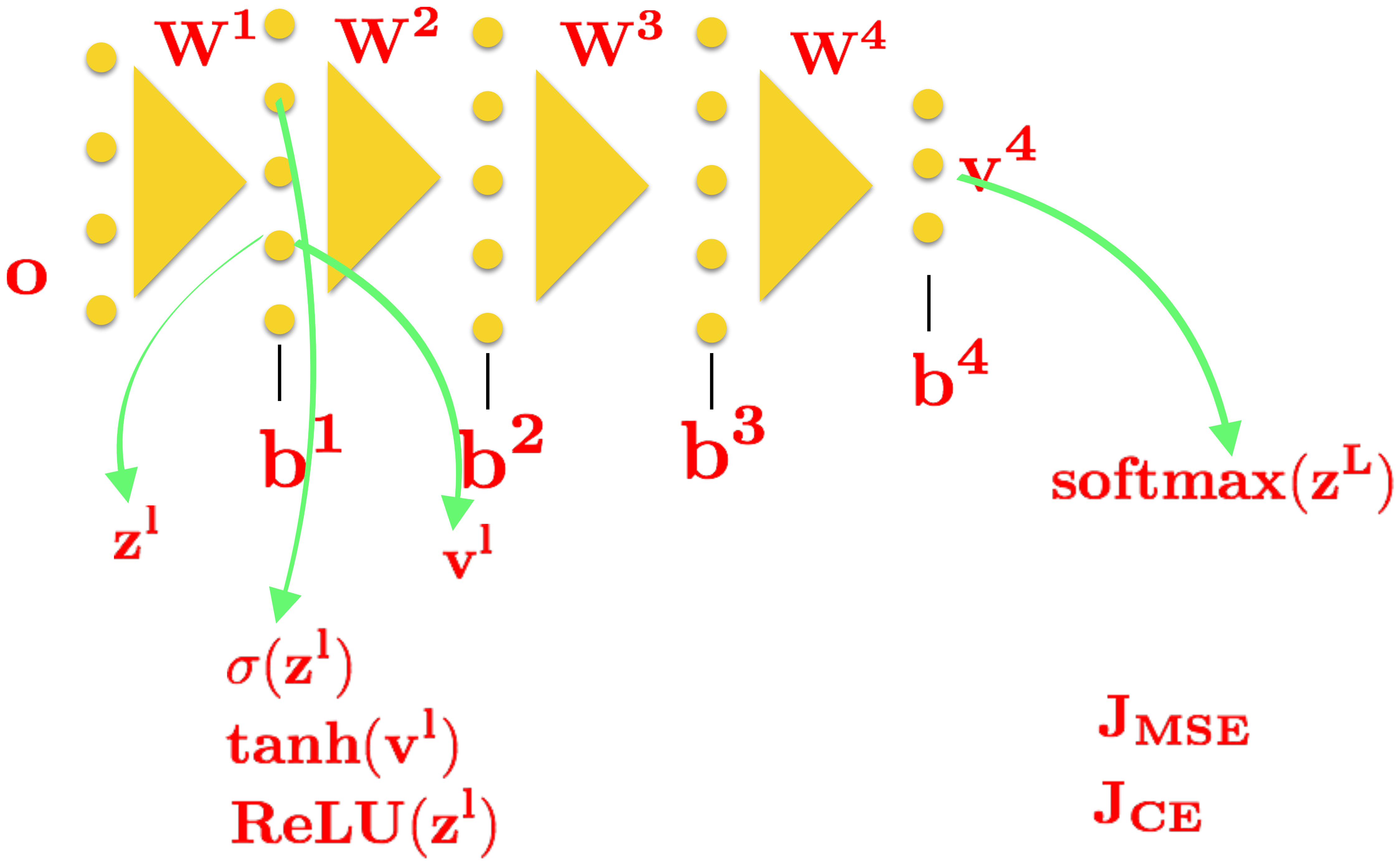
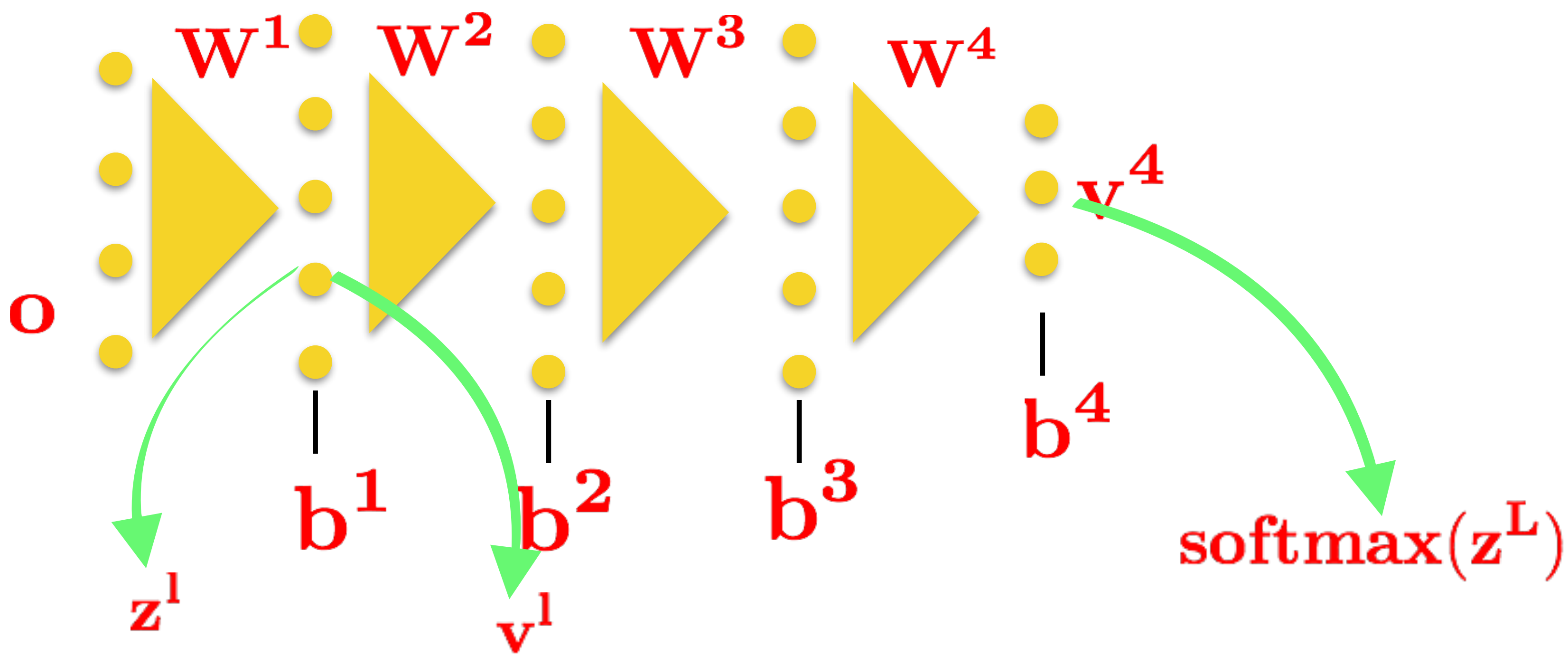
E9 261

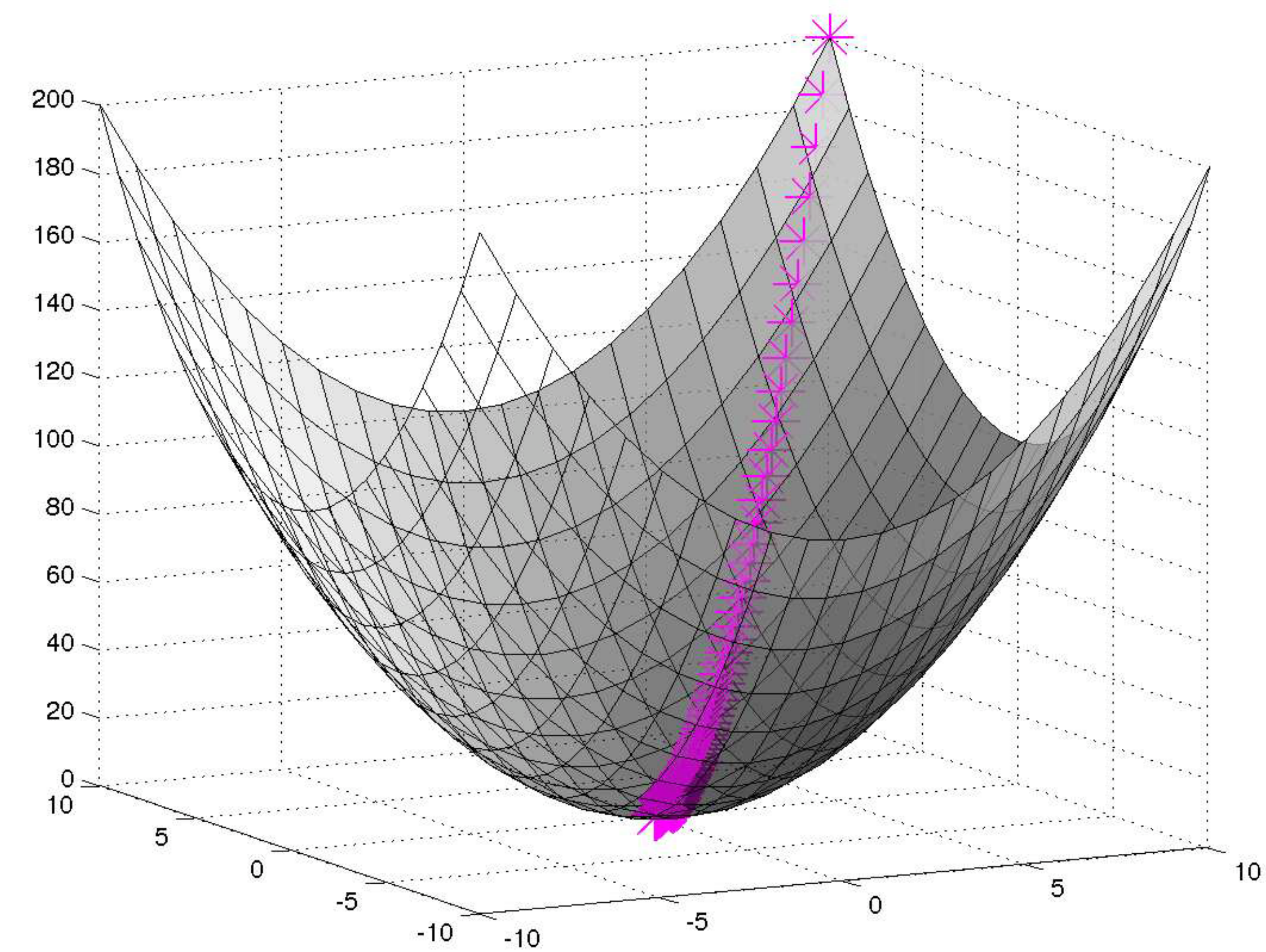
28-03-2016

DNN Definition



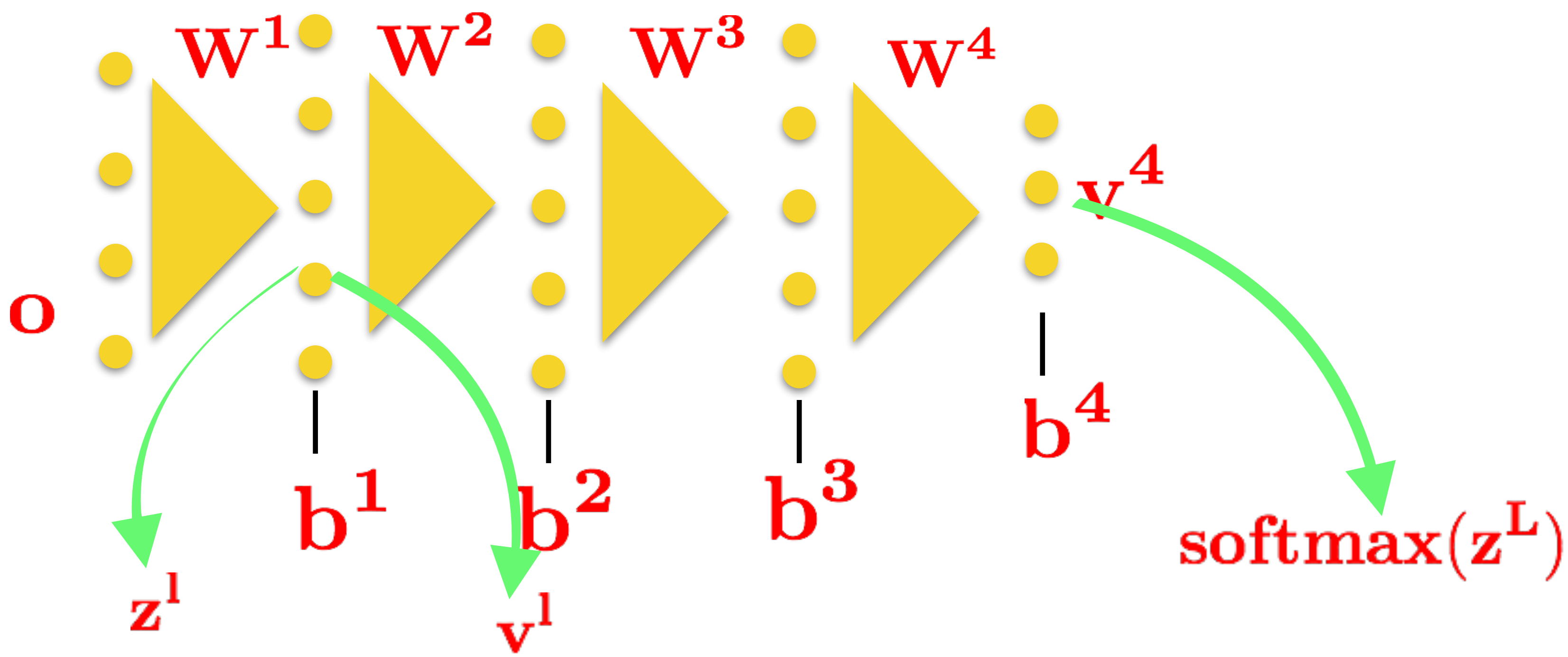
Back Propagation Learning





(Courtesy - D. Vijayasenan, NITK)

Back Propagation Learning



Parameter Update

$$\mathbf{W}^{l}_{t+1} = \mathbf{W}^{l}_{t+1} - \epsilon\,\Delta\mathbf{W}^{l}_{t}$$

$$\mathbf{b}^{l}_{t+1} = \mathbf{b}^{l}_{t+1} - \epsilon\,\Delta\mathbf{b}^{l}_{t}$$

Back Propagation Learning

Equivalence between MSE and CE based update

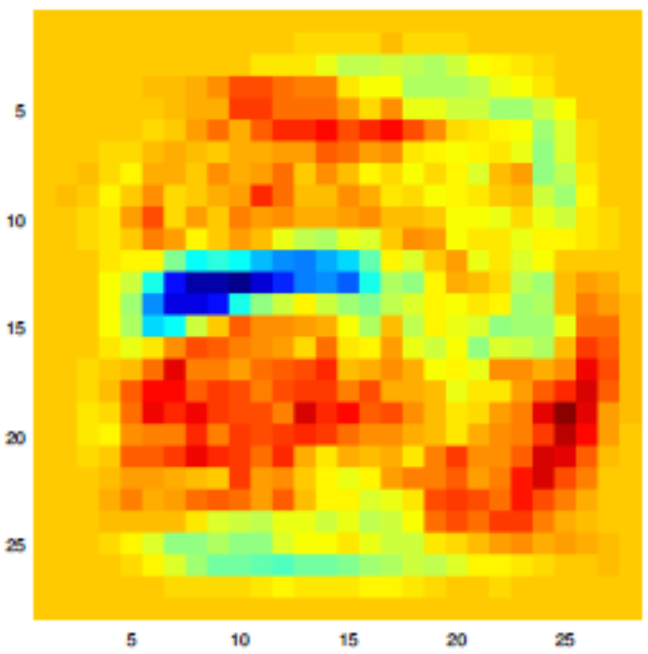
$$\Delta \mathbf{W}_t^L J_{MSE} = \Delta \mathbf{W}_t^L J_{CE} = (\mathbf{e}_t^L)(\mathbf{v}_t^{L-1})^T$$

$$\Delta \mathbf{b}_t^L J_{MSE} = \Delta \mathbf{b}_t^L J_{CE} = (\mathbf{e}_t^L)$$

$$\mathbf{e}_t^L = \mathbf{v}_t^L - y$$

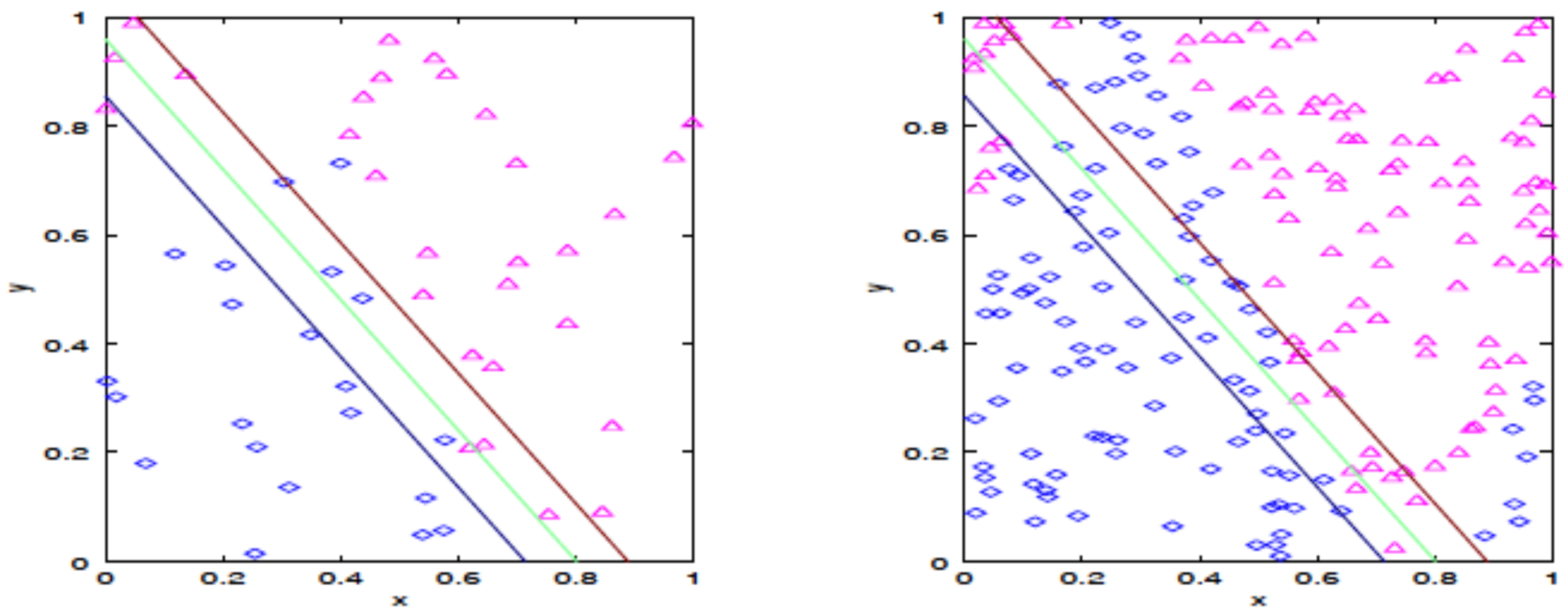
Back Propagation Learning

- detect the digit "2" from input image
- 28x28 input images
- 60000 samples
- Trained with a regularization parameter $\lambda = 10^-4$



(Courtesy - D. Vijayasenan, NITK)

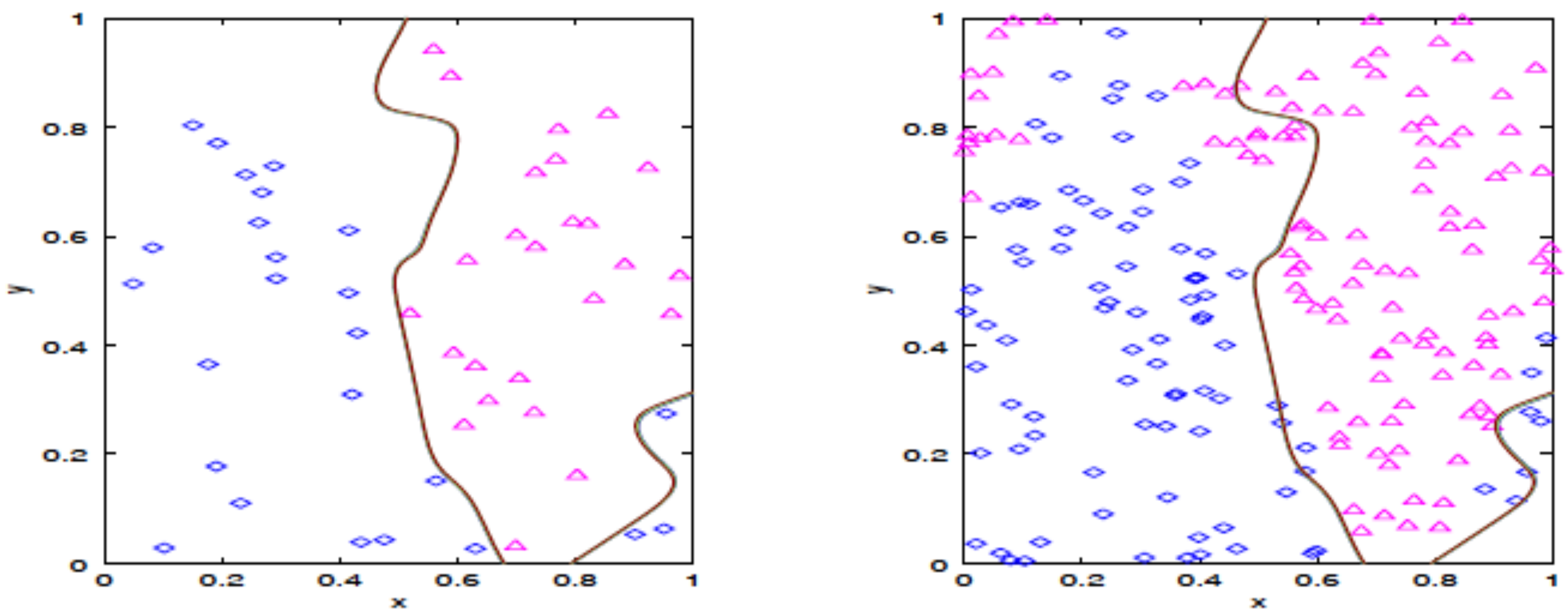
Underfit



- The model is not able to capture the variability in the data (Linear Model)
- Both the training and testing error are high (15%,20%)
- Try to learn a more complex model – more features, more hidden neurons, decrease regularization
- More data would not help

(Courtesy - D. Vijayasenan, NITK)

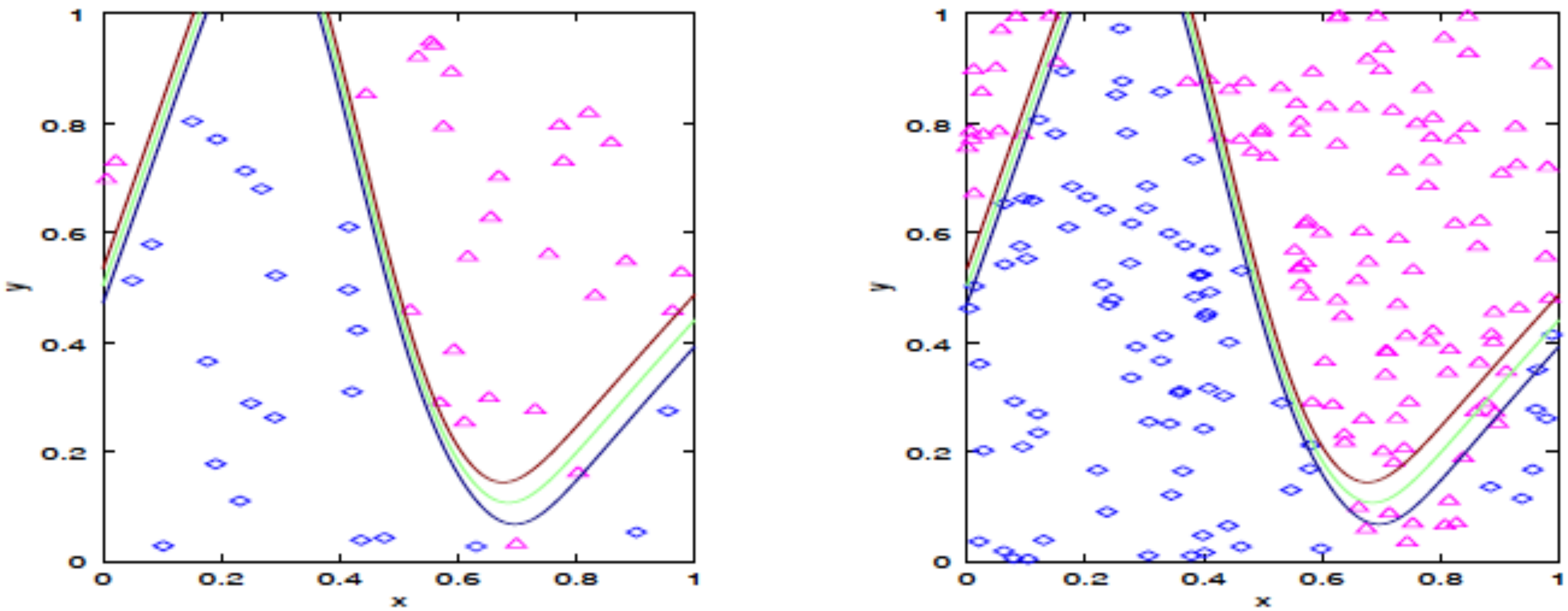
Overfit



- The model is capturing data as well as accidental variations (100 hidden neurons)
- Training error is too low and testing error is too high (0%, and 16%)
- Try to learn a simpler model – less features, less hidden neurons, increase regularization
- More data would help

(Courtesy - D. Vijayasenan, NITK)

Compromise



- Reasonable training and test errors – (4%, 8%)
- Appropriate model – capturing only the global characteristics not details

(Courtesy - D. Vijayasenan, NITK)

Other Considerations

- Data Preprocessing
- Model Initialization
- Overfit versus Underfit
  - Weight Decay
  - Dropout
- Batch Size Selection
- Pre-training
  - Restricted Boltzmann Machine