# ENGR 4350: Applied Deep Learning

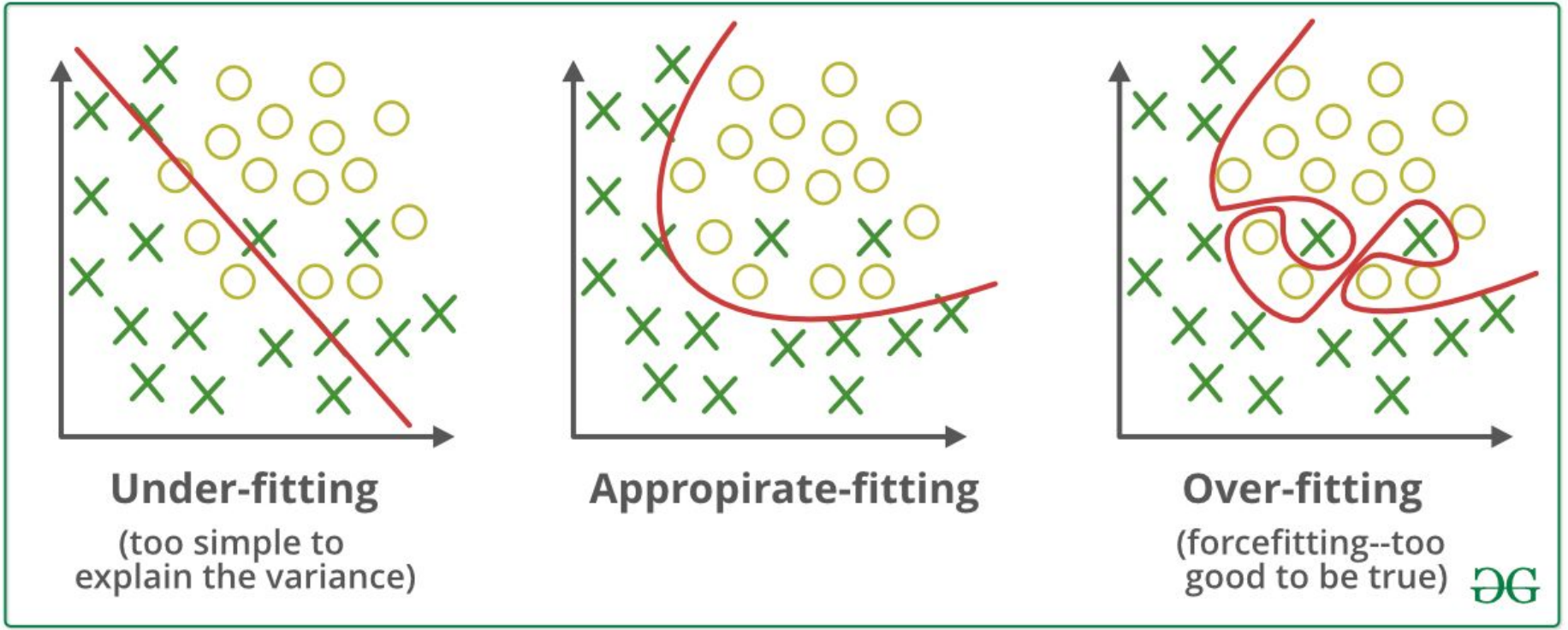
## Tuning

09/28/2022

# Outline

- Fitting Problem
- Hyperparameters
- Tuning Methods

# Under-Fitting & Over-Fitting



# Hyperparameters

- Learning rate
- Number of iterations
- Number of hidden layers
- Size of hidden layers
- Choice of activation functions
- (Choice of random seeds)

...

# Under-Fitting

- Learning rate
- Increase number of iterations
- Increase number of hidden layers
- Increase size of hidden layers
- Choice of activation functions
- (Choice of random seeds)

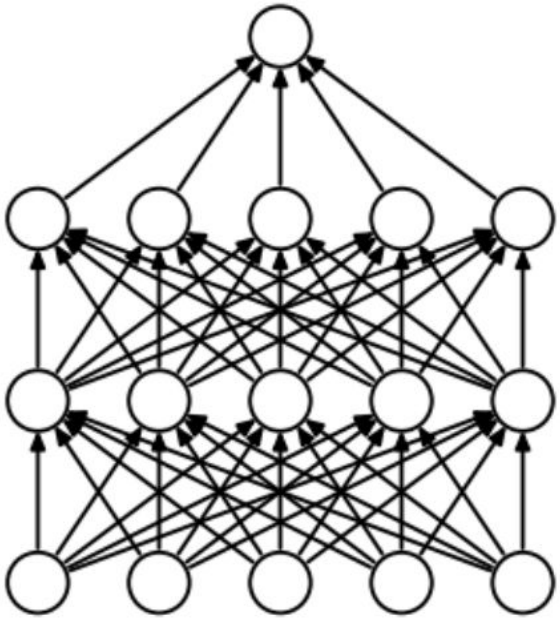
...

# Over-Fitting

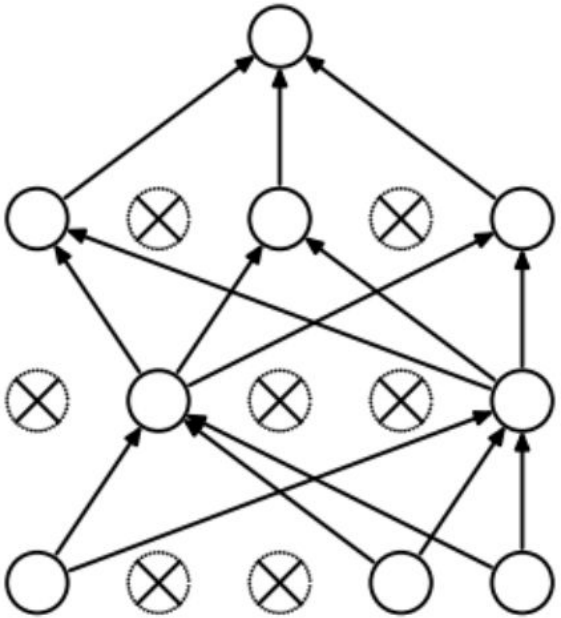
- Learning rate
- Decrease number of iterations
- Decrease number of hidden layers
- Decrease size of hidden layers
- Increase size of dataset
- Dropout regularization
- Data Augmentation
- Choice of activation functions
- (Choice of random seeds)

...

# Dropout Regularization



(a) Standard Neural Net



(b) After applying dropout.

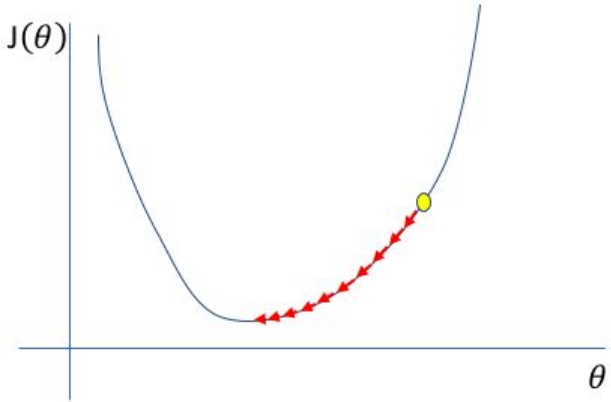
# Data Augmentation

# Training Convergence

- Learning rate
- Training data normalization
- Parameters initialization

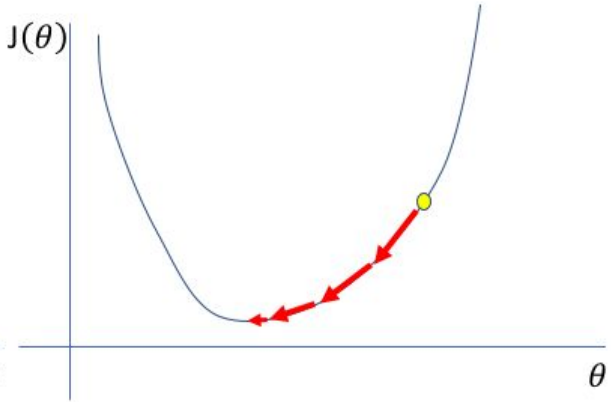
# Learning Rate

**Too low**



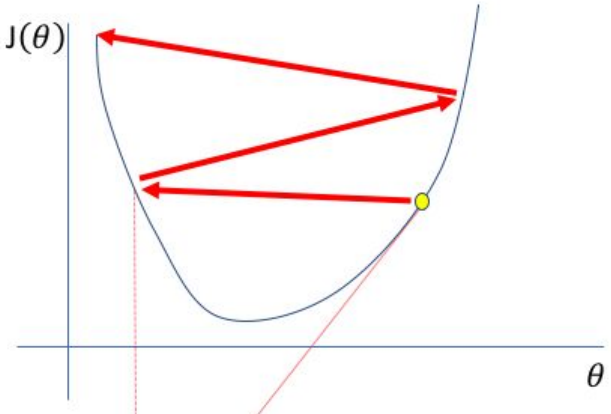

A small learning rate requires many updates before reaching the minimum point

**Just right**




The optimal learning rate swiftly reaches the minimum point

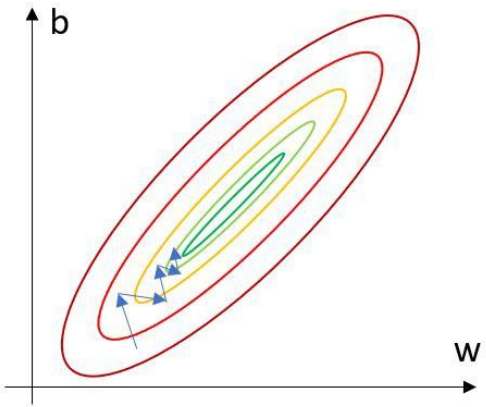
**Too high**



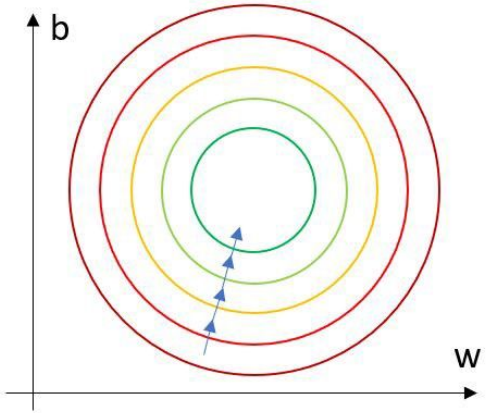

Too large of a learning rate causes drastic updates which lead to divergent behaviors

# Normalize Inputs

Unnormalized:




Normalized:




$$\mathbf{x} := \frac{\mathbf{x}-\mu}{\sigma}$$

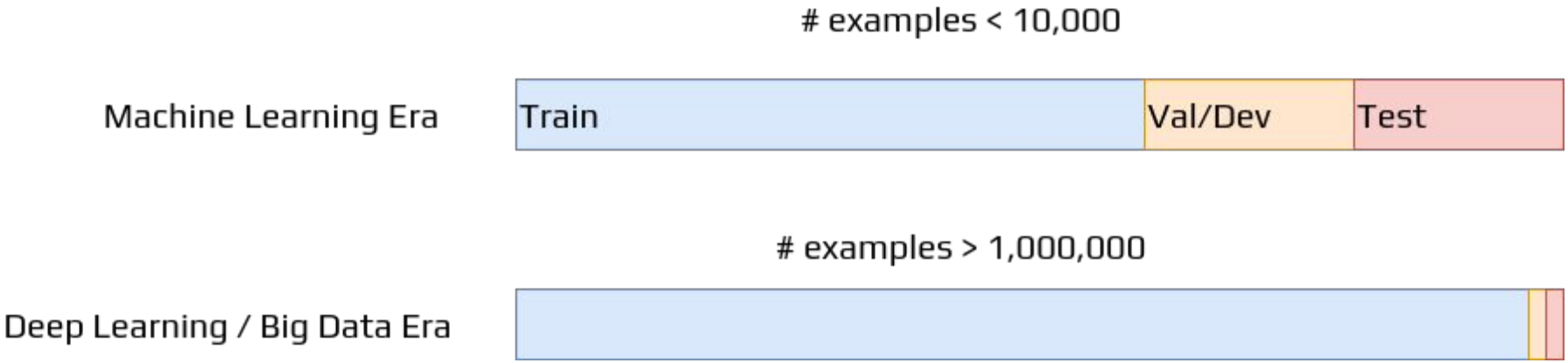
$$\mu = \bar{\mathbf{x}} = \frac{1}{M}\sum_{i=1}^{M} \mathbf{x_i} \qquad \sigma = \sqrt{\bar{\mathbf{x^2}} + \bar{\mathbf{x}}^{\mathbf{2}}} = \sqrt{\frac{1}{M}\sum_{i=1}^{M} \mathbf{x_i^2} + \left(\frac{1}{M}\sum_{i=1}^{M} \mathbf{x_i}\right)^2}$$

# Parameters Initialization

$$\mathbf{W} = \mathcal{N}(0,\, 0.1)$$

$$\mathbf{b} = 0$$

# Data Splitting



# Deep Learning is Alchemy