**Neural Networks**

**Course 2:**

**Hyper Parameters Tuning**

**Week 1**

**Submitted to Dr. Omar Nasr**

**Prepared by: Mona Mamdouh**

# **The Bias variance trade off**

The data is divided into three sets as follows

Test set

20%

Cross validation

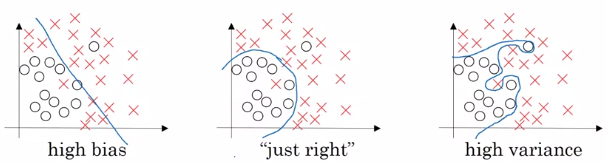
20%

Training set

60%

There is a tradeoff between the variance and the bias:

* If the bias is high it leads to under fitting the data to solve this we could use bigger network or train the network longer.
* If high variance is used it leads to over fitting the data to solve this we could use more data in training or use the regularization technique or create another NN architecture.



# **The regularization**

* Is used to solve the overfitting problem using three technique:

1. L2 regularization
2. Dropout regularization
3. Data augmentation
4. Early stopping

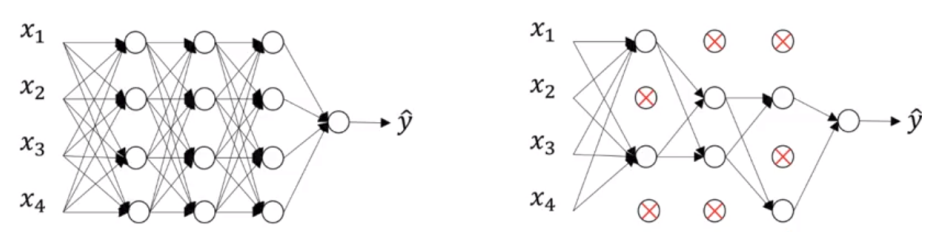
## **L2 regularization**

Where 𝜆 is the regularization parameter

## **Dropout regularization**

It is done by eliminating neurons to simplify the network as shown in figure 2.

It used during the training phase only and not used during the test phase as it add noise to the output.



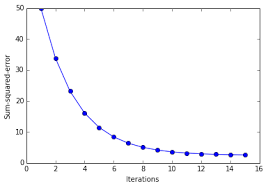
## **Data augmentation**

The data augmentation is used to increase the data set size by applying changes to the input data like mirroring, flip, rotate, zooming in and distortion to the training examples.



## **Early stopping**

It done by getting the number of iterations corresponding to the minimum cost value by drawing the curve as shown in figure 3 and get the best value of the number of iterations then compute the gradient decent.



# **Vanishing/exploding gradients**

One of the problems of training neural network, especially very deep neural networks, is data vanishing and exploding gradients. What that means is that when you're training a very deep network your derivatives or your slopes can sometimes get either very, very big or very, very small, maybe even exponentially small, and this makes training difficult. In this section you will see what this problem of exploding and vanishing gradients really means, as well as how you can use careful choices of the random weight initialization to significantly reduce this problem.