(Improving Deep Neural Networks Course)

Summary
On
Week 1

Name: Noha Ahmad Darwish
To: Dr/ Omar Nasr

- When training a NN, the greatest challenge will be the choice of the Hyperparameters like: Number of layers, Number of Hidden units, activation function and the learning rate.
- One more tickling challenge is the way the dataset is divided whether it's train/test data or train/dev/test data.
- The train/dev/test division has proved better results rather than train/test as the purpose of the dev data is the tuning of the hyperparameters. It can help us to decide on the network architecture or the optimization algorithm that would get us the best results on the test data.
- A very important thing to keep in mind while training a neural network is that the test data must come from the same distribution just like the training data.
- High Bias is when the network under fits the training data so the accuracy of the network on training data is small.
- To solve the bias problem we can: 1- enlarge the NN architecture
  2- Increase the training data maybe using data Augmentation
  3- enhancing the quality of the images, more clear
- High variance is when the network over fits the training data so the accuracy might be 99% on the training data while on the test set, the accuracy is 70%.
- To solve the variance problem we can: 1- decrease the size of the NN
  2- Increase the training data size
  3- use regularization
- Regularization is used to decrease overfitting.
- We have 3 types for regularization, L1, L2 and dropout.
- L1 and L2 Mainly used to force the network to keep weights as small as possible by adding the average weight value to the cost function so, as Weights increases >> cost function increases hence GD will try to minimize cost by minimizing the weight values.
- Dropout regularization is another approach also helps in increasing generalization and decreasing over-fitting
- the best way to improve generalization in decrease the over fitting is by increasing the data, which makes the network try to learn the most common feature between all training example which is the most useful features to generalize
- normalizing inputs have a large effect on increasing training speed as it make the data circular around the origin, which means we don't need a very small learning rate to avoid oscillation in one dimension
- Vanishing gradient occurs usually due to non-linear function saturation, for that reason the ReLu activation is preferable with normalizing the input and regularizing the weights.
- Exploding Gradients occurs due to multiplying gradient and accumulate them through deep-network which increase the value of gradient which in turn increase the weight update value and drive network to instability.

- To solve exploding gradient we may redesign a new network with less number of layers, or initialize weights with a more careful technique or using gradient clipping technique which clip the gradient value if it exceeded a certain limit.