Deep learning notes

What is a DNN?

* As researchers sometimes we are handed non-linear data, as nature can be unpredictable and a simple linear regression where we just draw a line of best fit will not help achieve suitable predictions based on the data being inputed .
* In light of data that is complex and has multiple patterns that occur through , it has been essential to devise of leverage new ways to deal with this data , this has come in the for of deep neural networks.
* These networks are comprised of a input layer that first deals with the data , a hidden layer which adds to the true depth within the model and is comprised of multiple neurons and multiple layers which serve the purpose of detecting and processing the different patters within the data and finally there is and output layer which provides the output required sometimes that can be a prediction like if a picture is a animal or human or it can be a linear prediction of what something like the stock market may look like in the coming days.
* Of course, like humans machines rarely get it right the first time and so usually these models are paired up with an optimizer function , these functions are require a loss metric which can reflect the accuracy of a prediction like how many times an network identified the correct animal in a picture which allows the system to adjust the parameters of the network accordingly by adjusting the weights associated with the neurons in the hopes that these new weights can help the model from a more accurate output(Alzubaidi *et al.*, 2021).
* More specifically this problem is referred to as the gradient problem which reflects how much the parameters must change (whether decreasing or increasing) to minimize the cost function. Each neuron has a local gradient calculated by the chain rule which gives each neuron within ever layer a local gradient of its partial derivatives of the outputs relative to its input. Finally to optimize these local gradients an algorithm often called gradient descent, uses adjusts the weights and biases of neurons which are initialized randomly but then are iteratively changed with the sole purpose of minimizing the cost function (Kostadinov, 2019).
* This algorithm also uses the learning rate, another pre-determined user defined number which decides how much influence the gradient has at each iteration, this number is normally kept small especially with large complex problems as to keep the model generalizable (can perform well on a new dataset) to a broad range of patterns across a large dataset(Wilson and Martinez, 2001).
* This iterative cost which minimizes the training process is referred to as epochs , the amount of iteration through the dataset(epochs) can also be chosen .Method like early stopping will be used to make sure the model is optimized sufficiently. Usually model users will have a training data to train the model , a small segment of data as validation and a final section to really test the effectiveness of the model called the test dataset. In early stopping the model is trained on the training data but is stopped when the loss function of validation data stops improving over a certain number of epochs this will make sure that the model is trained on the data but is also generalizable on new data by preventing the model from overfitting to the training data set(Bai *et al.*, 2021).