## TRAINING NEURAL NETWORKS

* Simple model is always preferred over a complex model as simple model might have a lesser training accuracy but it has a larger testing accuracy than the complex model.
* Small values of epoch have high training and testing errors,the optimal epoch value has small training and testing errors ,and epochs> optimal epoch has a very less training error and a slightly high testing error.
* Models with high coefficients tends to have lower errors but may also overfit so,

L1 Error function = (-1/m)\* + Lambda\* (|w1| + |w2| +…|wn|)

L2 Error function = (-1/m)\* + Lambda\* (w1^2 + w2^2 +…wn^2)

* L1 regularization is better for feature selection while L2 regularization is better for training models
* Dropout – we drop or disable some perceptrons from the training set while training the neural network
* Gradient descent won’t work in complex sets which has a series of ups and downs so random restart- a method where the training is done at various starting points is done to identify the global minima
* Another patch is to change the activation function:

Tanh(x) =( e^x – e^-x)/(e^x + e^-x)

Relu{Rectified Linear Unit} = x if x>= 0

= 0 if x<0

* High learning rates – fast models but fails to capture the minima hence inaccurate

Low learning rates – slow models but captures the minima hence accurate enough

* Momentum – While using gradient descent we get stuck on a local minima so to go over dt we use a concept called momentum i.e we add the sum of products of the previous steps with powers of momentum

Step(n) += Beta\* Step(n-1) +Beta^2(Step(n-2))…