Inzva DLSG HW1 Answers Celaleddin Ömer Sağlam:

1. Deep Learning is Machine Learnings subarea of neural networks. It is “deep” because you need deep structures of neural networks to make model learn something, wide learns aswell but requires exponentially growing sizes so it is not efficient. Making it deep is efficient on the other hand. Learning is because we are using backpropagation to alter variables and maket hem represent some features in some manner.

In Traditional ML, we design features but DL learns features itself. DL has much more need for bigger datasets, DL is somehow more complex and uncomprehensible than TML. DL generally Works better when needs are satisfied but opportunity cost is computation time and force. For smaller datasets, TML may give better results.

1. Forward Prop means, we put an input and multiply them with filters(fully connected layers etc. are also a filters of some kind. They transform our data into some other format. Linear or convolutional etc.) İt until we reach a decision point. Than we see what is the result. Process is, on basis of math, matrix multiplication by hard.
2. Back Prop is the last point humanity get that resembles human brain. We make machine see something, compare it to a higher order knowledge(labels) and make it change itself to fit that higher knowledge. Basically what we do as humans in a nutshell. We see the World as a child. Elders, teachers, books etc say us something that is true and we learn how to thing in that way. Learning is via changing parameters. We don’t know how neurons learn parameters actually but artificial neurons learns via gradients.
3. Activation Functions are essentially decision points. It makes process different than linear transformations and make us represent complex functions. You can see it as bending the space or deciding importance of weights etc. Because of this decision propertly, we can make NN’s learn more and more complex structures.
4. A1:6\*1 W1:6\*4 b1: 6\*1 Z1:6\*1

A2:4\*1 W2:4\*6 b2:4\*1 Z2:4\*1

A3:3\*1 W3:3\*4 b3:3\*1 Z3:3\*1

Z(l) = W(l)\*a(l-1) + b(l)

A(l) = activation(z(l))

1. İf you dont train enough or dont have big enough model; you will underfit.

If you train too much, dont do proper regularization, model is big enough, have too less data, have biased data; You will overfit.

High train loss means underfit. Too low train loss with huge gap of val and test loss/accuracy means overfit.

1. LR means size of step. You go towards a minimum point. Your step size is LR. If you have too small steps, you wont reach there in time or you wont pas seven smallest local minima points. Too high than you are not precise enough to go down. You will diverge or oscillate.
2. Force you to budget your decisions. It says you can’t trust on the input you see always because i will take random part of them and make you decide without them. So you have to learn your inputs instead of memorizing. Some kind of regularization effect that Works via cutting your arms of temporarily and seeing you trying to do same thing as before.
3. You can change your friends but your family. You learn weights and optimize them while living your life. But once you are born, you have to stuck with your hyperparameter family. But the creator may give you alternative universes. Starts your life with different families and see how well you decide your friends.
4. BGD updates gradients according to whole dataset. SGD makes it change for every sample. MBGD is middle ground. Put them into small groups and give them autonomy.
5. Adam is combination of RMSprop and Momentum. Momentum keep you on line even when small distractions happen. RMSprop makes you smarter about your step size. If you are on highway, you go 300km/h. But when you got near your house, you go down to 30 so you can see your wife and wave them while passing. Bigger gradients with small LR, small gradients with large LR. Also helps you to stay on average. Averages out gradients among parameters.

BYE ☺