**DL theory : Assingments-14**

1. It is generally not okay to initialize all the weights to the same value, even if that value is selected randomly using He initialization. This is because it can lead to a phenomenon called "symmetry breaking", where all neurons in the same layer will learn the same feature and the model will not be able to learn any useful representations.
2. It is okay to initialize the bias terms to 0, as long as the weights are initialized with a different value. The bias terms are typically initialized to 0 because they do not affect the symmetry breaking issue as much as weight terms.
3. Three advantages of the ELU activation function over ReLU are that it can produce sparse models, it can handle negative input values, and it has a mean activation closer to 0, which can help to reduce internal covariate shift.
4. ELU is generally used when the data has a Gaussian distribution, LeakyReLU variants are used to overcome the "dying ReLU" problem, ReLU is often used in feedforward neural networks, Tanh is often used in Recurrent Neural Networks (RNNs) and LSTMs, Logistic is often used in logistic regression and binary classification problems, and Softmax is often used in multi-class classification problems.
5. If the momentum hyperparameter is set too close to 1, the optimizer may overshoot the minimum of the cost function and oscillate around it, which can lead to slow convergence or even diverging.
6. Three ways to produce a sparse model include using L1 regularization, using sparse activation functions such as ReLU, and using dropout.
7. Dropout does not slow down training, but it does slow down inference. This is because during training, dropout is used to randomly drop out some neurons in the network, which makes the model more robust. However, during inference, all neurons are used, which makes the predictions slower.