Q1: Is it okay to initialize all the weights to the same value as long as that value is selected randomly using He initialization?

Ans: No, initializing all weights to the same value, even if using He initialization, can lead to symmetry between neurons, which can cause them to learn identical features and limit the model's capacity to learn.

Q2: Is it okay to initialize the bias terms to 0?

Ans: Yes, initializing bias terms to 0 is a common practice and often works well in practice.

Q3: Name three advantages of the ELU activation function over ReLU.

Ans: The advantages of the ELU activation function over ReLU are:

Smoothness, which reduces the likelihood of "dying ReLU" problem and can accelerate convergence.

Negative values for negative inputs, which allows neurons to have negative outputs, making them more expressive and better able to learn complex functions.

Better performance on tasks involving image recognition and object detection.

Q4: In which cases would you want to use each of the following activation functions: ELU, leaky ReLU (and its variants), ReLU, tanh, logistic, and softmax?

Ans:

ELU: Good default choice for most tasks, especially image recognition and object detection.

Leaky ReLU (and its variants): Useful when there is concern about dying ReLU problem, but not as good as ELU in general.

ReLU: Good default choice if computational efficiency is a concern, but not as good as ELU for many tasks.

tanh: Useful for output layers of binary classifiers and some recurrent neural networks (RNNs).

logistic: Useful for output layers of binary classifiers, especially when the outputs need to be interpretable as probabilities.

softmax: Useful for output layers of multiclass classifiers, especially when the outputs need to be interpretable as probabilities.

Q5: What may happen if you set the momentum hyperparameter too close to 1 (e.g., 0.99999) when using a MomentumOptimizer?

Ans: If the momentum hyperparameter is set too close to 1, the optimizer will have very little capacity to change direction and may overshoot the minimum of the cost function, leading to instability and poor performance.

Q6: Name three ways you can produce a sparse model.

Ans: Three ways to produce a sparse model are:

L1 regularization, which encourages the weights to be close to 0, resulting in many of them being exactly 0.

Dropout, which randomly sets some neurons to 0 during training, effectively removing them and creating a sparse network.

Explicit pruning, which involves removing weights or entire neurons that have little effect on the output of the network.

Q7: Does dropout slow down training? Does it slow down inference (i.e., making predictions on new instances)?

Ans: Dropout can slow down training, as the network needs to train more neurons to compensate for the dropped-out ones. However, it usually results in better generalization and can ultimately reduce the number of training iterations needed. Dropout does not slow down inference, as all neurons are used during inference and no dropout is applied.