A1

(a)

False. Deep neural networks have non-convex loss surfaces, so gradient descent does not guarantee the global optimum.

(b)

False. Initializing all weights to zero prevents breaking symmetry, causing identical updates and hindering training.

(c)

True. Non-linear activation functions enable the network to learn non-linear decision boundaries, which would be impossible with purely linear transformations.

(d)

False. Although the backward pass is more expensive than the forward pass, it is typically of the same order of magnitude and not prohibitively larger (big O time is the same).

(e)

False. Neural networks are powerful and extensible, but they are not always the best choice for every circumstance due to factors like data requirements, computational cost, and interpretability.