**CONCLUSIONS AND OBSERVATIONS**

**1. Learning Rate**

* A high rate (0.1) led to rapid initial decrease but unstable oscillations/divergence.
* A medium rate (≈0.01) provided smooth convergence and ~90% validation accuracy.
* A low rate (0.001) was stable but too slow; validation accuracy lagged by epoch 10.

**2. Batch Size**

* Batch=1 (online) produced very noisy but sometimes slightly better final accuracy due to regularizing noise.
* Batch=32 balanced stable gradients with update frequency—our recommended default.
* Batch=256 yielded smooth trajectories but ~1% lower final accuracy and slower generalization.

**3. Number & Width of Hidden Layers**

* 0 layers (linear model) peaked at ~84% validation accuracy—underfitting.
* 1 hidden layer of 128 neurons boosted accuracy to ~88%; non-linearity is beneficial.
* 2 hidden layers (256→128) reached ~90% but added ~30% training time; diminishing returns.

**4. Loss Function**

* Cross-Entropy: fastest convergence and highest final accuracy.
* MSE: slower training, less stable gradients, plateauing ~86%.
* MAE: very slow and erratic—unsuitable for classification.

**5. Over-fitting**

* Deeper or wider models hit ~100% train accuracy by epoch 10 while validation accuracy plateaued or dipped; use early stopping or regularization.

**Recommended Best Setup**

* Learning rate = 0.01; batch size = 32; 2 hidden layers (256→128); Cross-Entropy loss; 8–10 epochs with early stopping.