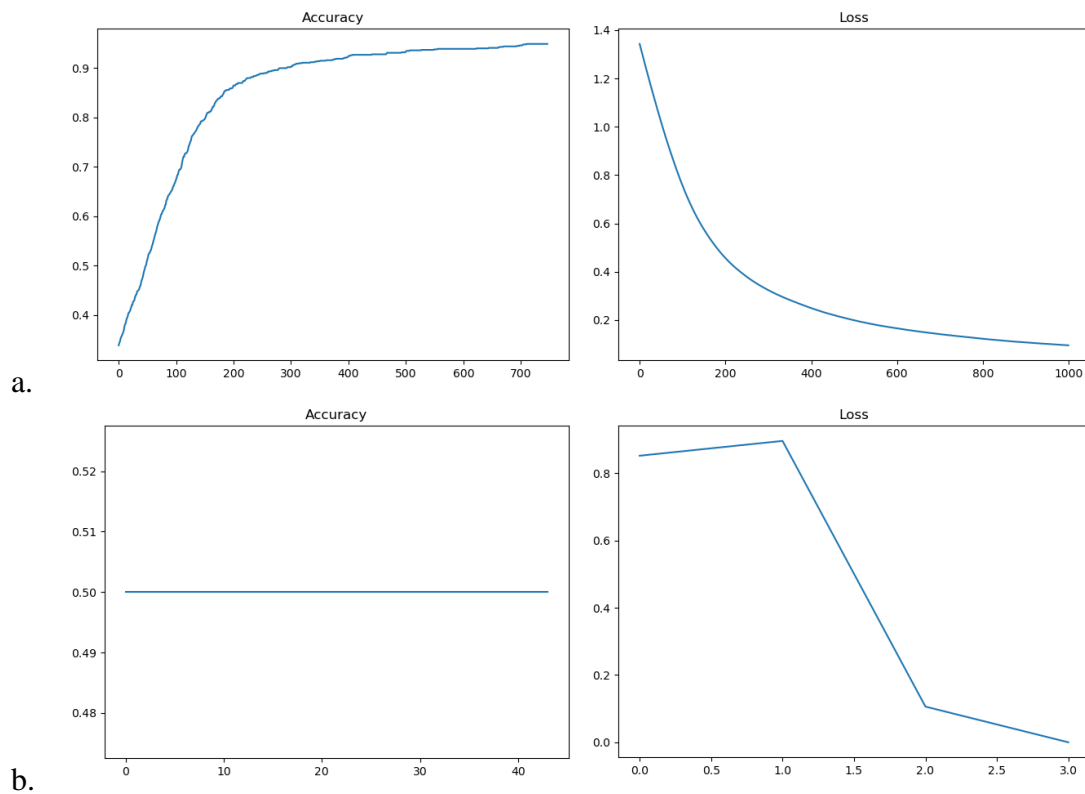1. Visualizing Gradient Descent:

a.

b.

????

c. Stochastic gradient descent was much faster to train. The loss jumps back and forth in stochastic gradient descent while it continuously decreases in batch.