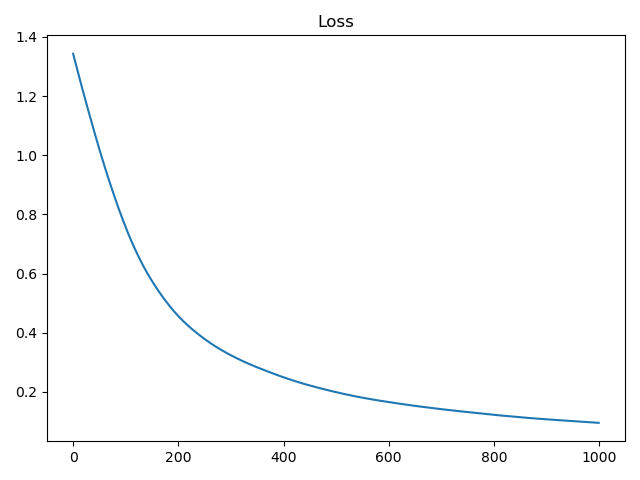
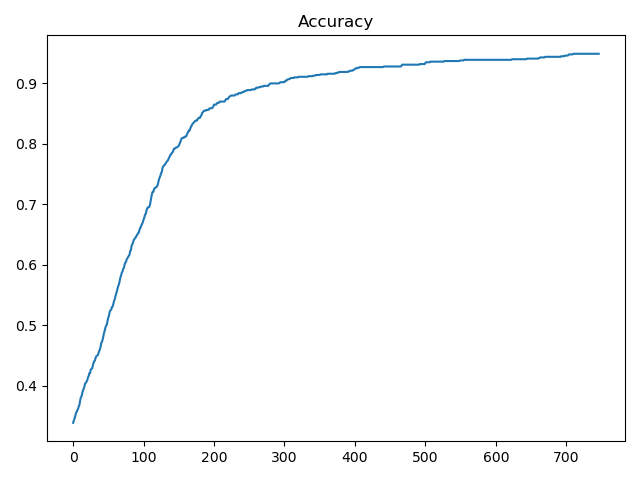
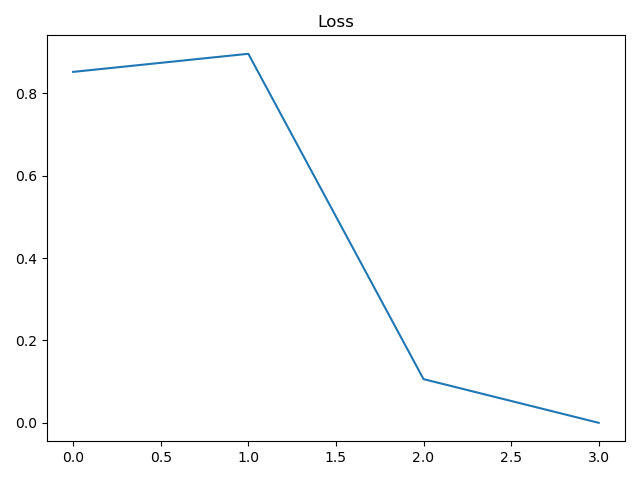
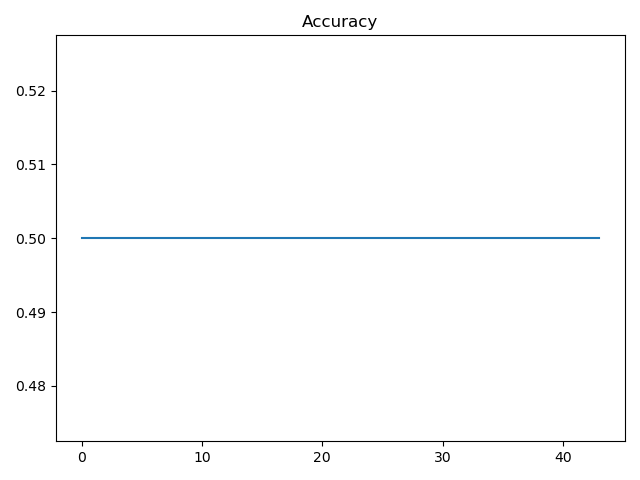
1. Visualizing Gradient Descent:
   1. 
   2. ????
   3. Stochastic gradient descent was much faster to train. The loss jumps back and forth in stochastic gradient descent while it continuously decreases in batch.