Obstacle 1: Initially, my sampling function was very slow, taking over a minute to run. I realized I was accidentally copying the entire 512×512 image on each iteration using an unnecessary intermediate variable.

Obstacle 2: Dimension mismatch in forward pass. When computing `z2 = data.dot(W1) + b1`, I got a broadcasting error because `b1` had shape `(25, 1)` while `data.dot(W1)` had shape `(10000, 25)`. I used `print(data.shape, W1.shape, b1.shape)` to identify the issue. I transposed `b1` using `b1.T` so it broadcasts correctly as `(1, 25)` across all training examples.

Obstacle 3: Debugging the cost function components. Following the README's debugging tip [2], I first set `lambda = beta = 0` and implemented only the squared error term. I verified this worked with gradient checking before adding the weight decay and sparsity penalty terms. This made it much easier to isolate bugs.