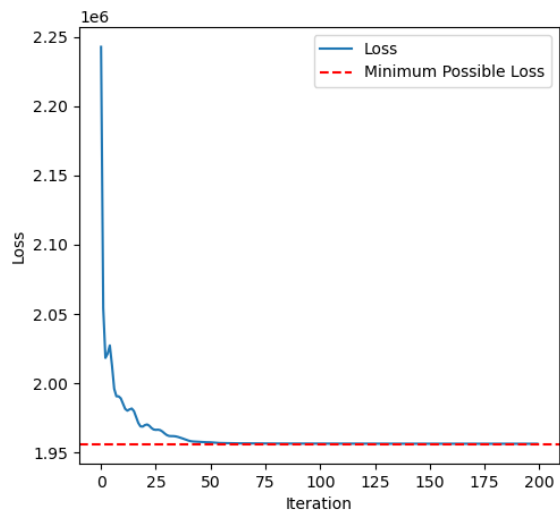1. The loss as a function of 200 iteration, the learning rate being 1.

2. The final token probabilities compared to the (known) optimal probabilities.