Daniela Jiménez Lara. Natural Language Processing.

Assignment 2.

Gradient descent

By selecting 60 iterations and a learning rate of .20, the following graphs illustrate the results.

A graph with a line

Description automatically generated

A graph with numbers and letters

Description automatically generated

A graph with green dots

Description automatically generated

But by augmenting the iteration to 100 and keeping the learning rate to 0.02, the difference between the token probabilities and optimal probabilities decreases.

A graph with a line

Description automatically generated

A graph with numbers and letters

Description automatically generated

A graph with green dots

Description automatically generated