Comparing the train losses for the three models Full model One dropout layer Two dropout layers 2.0 Neg. log likelihood loss 1.5 1.0 0.5 0.0 100000 200000 300000 400000 Training examples the model has seen