Flow

1. Fix seed to 42
2. Create labeled train data
3. Generate vocabulary – train words and two preceeding words, and two following words for end of sequence
4. Construct computation graph according to the specifications
5. Train
   1. Between each train iteration check dev accuracy, calculated according to the specifications
   2. With word that is in train data but not dev data –
      1. Give majority label - Naïve
      2. Randomly initialized vector – uses learned embeddings
      3. Search for same windows in train data and takes majority label – some or all sub combinations in 4 window (prior two, prior 1, following 1, following 2, prior two and following 1, all 4 and so on) – most complex but probably works good.

﻿Experiment with several network configurations, learning rates. Generate the graphs requested.