



Input layer w and output layer y have the same dimensionality as the vocabulary (10K - 200K)

Hidden layer s is orders of magnitude smaller (50 - 1000 neurons)

U is the matrix of weights between input and hidden layer, V is the matrix of weights between hidden and output layer

Without the recurrent weights W , this model would be a bigram neural network language model