

Dependencies on input

Trick: Pretend the current input x_t contains not only the current word w_t , but also w_{t-1} and w_{t+1} and build observation functions for them as well.

The model is discriminative, so we can use the whole input:

$$[y_t = y][y_{t-1} = y']$$

