Sequence Labeling, Tagging with Classifiers

Input $\bar{x} = (x_1, x_2, \ldots, x_n)$

Output $\bar{y} = (y_1, y_2, \ldots, y_n)$ one prediction per word

Structured Classification

Ex. Fed raises interest rate 0.5 percent

- Predict each y_i independently w/ logistic regression

$$P(y_i = y \mid \bar{x}, i)$$ ← index we're predicting at

- BOW: $f(\bar{x}, y = NN) = [\ 000\,f(\bar{x})\,000 \mid 100100 \mid 0000 \ldots\ ]$ (100100 under NN)

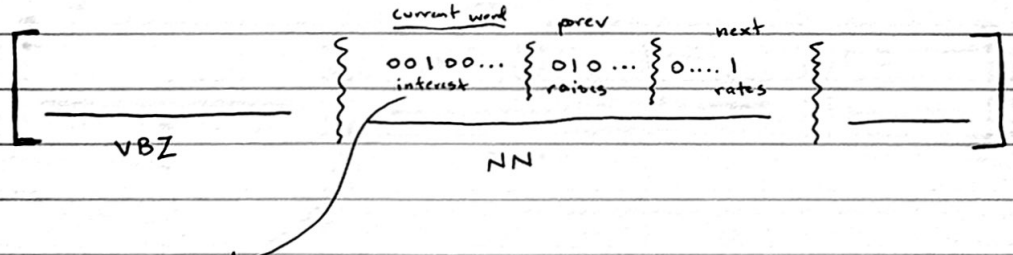
* This DOES NOT work b/c independent of i, doesn't look at word being tagged

- Positional: $f(\bar{x}, y = NN, i = 3)$ * Simple way: single feature on current word

$[\ 0\ldots \mid 00100000 \mid 0\ldots\ ]$ (00100000 = interest, under NN)

- Positional Features w/ context

$f(\bar{x}, y = NN, i = 3)$



Conjunction of several properties

Indicator [curr word = interest ∧ tag = NN]

treat as "word" in bag-of-words space

- Problem w/ Classification for tagging

indicators → classifier

* predictions of classifier may be incoherent

What goes wrong? : Not making use of output structure

VBZ ≠ VBP

NNS ≠ NN

Fed raises interest rates ...