

Review of Nov 12

Generative Models for classification

Bayes decision rule: $y = \underset{y \in \mathcal{Y}}{\operatorname{argmax}} \{ P(x=x | y=y) \cdot P(y=y) \}$

Prior: $P(y=y)$

Class conditional distributions: $P(x=x | y=y)$

Multivariate Naive Bayes: $P(x=x | y=y) = \prod_{i=1}^N P(x_i=x_i | y=y)$

Multinomial Naive Bayes: $P(x=x | y=y) = \prod_{i=1}^{k+1} P(w=x_i | y=y)$

Linear Discriminant Analysis: $P(x=x | y=y) = N(\mu_y, \Sigma)$

Part-of-Speech Tagging

Instance space X : Sequences of words

Target space Y : Sequences of POS tags

→ sequence prediction

Bayes decision rule: $y = \underset{y \in \mathcal{Y}}{\operatorname{argmax}} \{ P(x=x | y=y) \cdot P(y=y) \}$

$P(y=y)$ (Prior): $P(y=(y_1, \dots, y_e))$

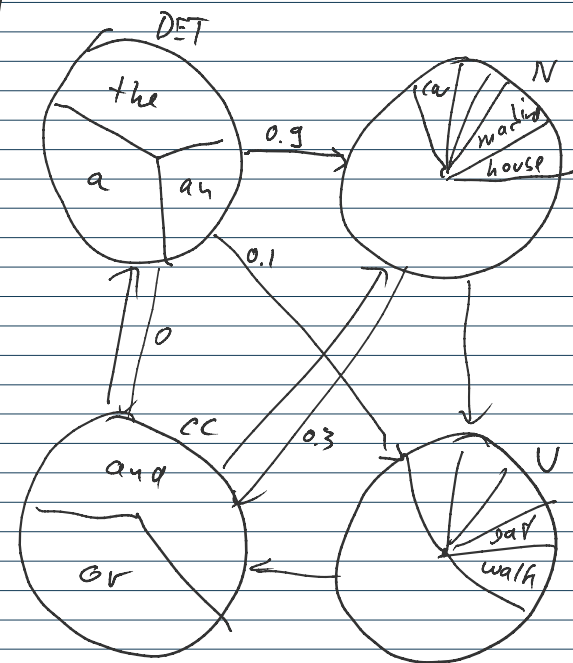
→ Assume dependencies follow Markov model

$$P(y=(y_1, \dots, y_e)) = P(y_1=y_1) \cdot \prod_{i=2}^e P(y_i=y_i | y_{i-1}=y_{i-1})$$

$P(x=x | y=y)$: Class conditional

→ Assume the word x_i is independent of other words and all tags except y_i .

$$P(x=(x_1, \dots, x_e) | y=(y_1, \dots, y_e)) = \prod_{i=1}^e P(x_i=x_i | y_i=y_i)$$



x the house and
y DET N CC

Example:

$$P(x=(I, \text{bank}, \text{at}, \text{CFCU}), y=(N, V, N, V))$$

$$= P(y_1=N) \cdot P(x_1=I | y_1=N)$$

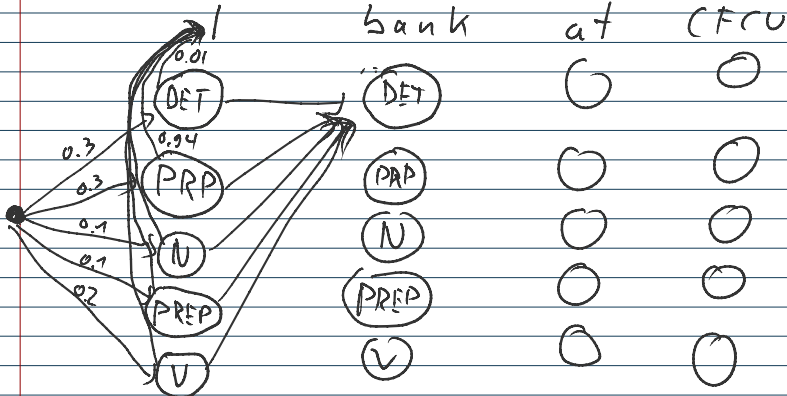
$$\cdot P(y_2=V | y_1=N) \cdot P(x_2=\text{bank} | y_2=V)$$

$$\cdot P(y_3=N | y_2=V) \cdot P(x_3=\text{at} | y_3=N)$$

$$\cdot P(y_4=V | y_3=N) \cdot P(x_4=\text{CFCU} | y_4=V)$$

$$= 0.1 \cdot 0.01 \cdot 0.19 \cdot 0.4 \cdot 0.3 \cdot 0.01 \cdot 0.19 \cdot 0.01 = 4.332 \cdot 10^{-10}$$

Viterbi Example



$$S_{DET}^{(1)} = 0.3 \cdot 0.01$$

$$S_{PRP}^{(1)} = 0.3 \cdot 0.94$$

$$S_N^{(1)} = 0.1 \cdot 0.01$$

$$S_{PREP}^{(1)} = 0.1 \cdot 0.01$$

$$S_{DET}^{(2)} = \max \left\{ \begin{array}{l} DET \ 0.003 \cdot 0.01 \cdot 0.01 \\ PRP \ 0.282 \cdot 0.01 \cdot 0.01 \\ N \ 0.001 \cdot 0.01 \cdot 0.01 \\ PREP \ 0.001 \cdot 0.3 \cdot 0.01 \\ V \ 0.002 \cdot 0.2 \cdot 0.01 \end{array} \right\} = 2.82 \cdot 10^{-5} \text{ PRP}$$

S	1	2	3	4
DET	0.003	$2.82 \cdot 10^{-5}$		
PRP	0.282	$2.82 \cdot 10^{-5}$		
N	0.001	$1.13 \cdot 10^{-3}$		
PREP	0.001	$5.64 \cdot 10^{-4}$	$1.25 \cdot 10^{-2}$	$1.5 \cdot 10^{-3}$
V	0.002	$2.61 \cdot 10^{-2}$		

Time complexity

~ linear in length of sequence

~ cardinality of tag set \rightarrow quadratic