

Review | Q & A Session

Statistical Models :

- Generative models $\begin{cases} \text{Naive Bayes Model} \\ \text{Markov Models of } k\text{-th order} \\ \text{HMM (Hidden Markov Model)} \end{cases}$
- Discriminative models : Perceptron (Log-Linear Model)

1. Naive Bayes Model

$$P(y, x_1, \dots, x_m) = P(y) \prod_{i=1}^m P(x_i | y)$$

$$\text{Log-Prob. form: } \log P(y, x_1, \dots, x_m) = \log P(y) + \sum_{i=1}^m \log P(x_i | y)$$

$$X = x_1, x_2, \dots, x_m \quad \begin{array}{l} m: \text{length of text.} \\ y: \text{label} \end{array}$$

Application: Text classification.

2. Markov Model of k -th order

$$P(x_1, x_2, \dots, x_m) = \prod_{i=1}^{m+1} P(x_i | x_{i-k}, \dots, x_{i-1})$$

2nd- Markov model

$$P(x_i | x_{i-2}, x_{i-1})$$

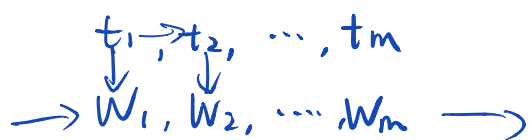
k : order

m : len of seq.

i : pos.

App.: Language modeling (predicting the probability of a given text (sequence of words))

3. HMM. of k th order



$$P(t_1, \dots, t_m, w_1, \dots, w_m) = \left[\prod_{i=1}^m P(y_i | y_{i-k}, \dots, y_{i-1}) P(w_i | t_i) \right]$$

Application: Part-of-Speech tagging

4. Perceptron (Log-linear model)

$$P(y|x) = \frac{1}{N(x)} e^{\sum_k W_k m_k(x, y)}$$

y : text label
 x : word sequence

$N(x)$: normalization of text

W_k : weight (model parameter)

m_k : feature value (e.g. word freq. in NP).

Application: Text classification