Review | Q & A Session

Statistical Models:

- Generative models
  - Naive Bayes Model
  - Markov Models of k-th order
  - HMM (Hidden Markov Model)
- Discrimitive models: Perceptron (Log-Linear Model)

1. Naive Bayes Model

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

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

$X = x_1, x_2, \cdots, x_m$ m: length of text.
y: Label

Application: Text classification.

2. Markov Model of k-th order

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

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 kth order

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

Application: Part-of-Speech tagging

4. Perceptron (Log-Linear model)

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

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 NLP).

Application: Text classification