Graphical Models

MMs and HMMs

St. Joseph's University, Bengaluru

ML 2 BDA3321

Markov Models

Markov chain

We assume discrete time steps, the **Markov Chain** as follows:

$$p(X_{1:T}) = p(X_1) \prod_{t=2}^{T} p(X_t | X_{t-1})$$

Markov Models

Transition Matrix

A **Transition matrix** is a $K \times K$ matrix, where $A_{ij} = p(X_t = j | X_{t-1} = i)$ is the probability of going from state i to state j, and $\sum_j A_{ij} = 1$. We assume X_t is discrete so $X_t \in \{1, \ldots, K\}$

State Transition Diagram

A graphical representation of the transition matrix is known as the **state transition diagram**. The weights associated with the arcs are the probabilities.

Markov Models

Applications: Language Modeling

Uses of Language models:

- Sentence Completeion
- Data Compression
- Text Classification
- Automatic essay writing

Markov Models

Application: Language Modeling

In general, we can build an n-gram model.
The marginal probalities guven by $p(X_t = k)$ are called the **unigram statistics**.

Language Modeling

MLE for Markov language models

- The probability of any particular sequence of length T is given by

$$p(x_{1:T|\theta}) = \pi(x_1)A(x_1, x_2)\dots A(x_{T-1}, , x_T)$$
$$= \prod_{j=1}^{K}(\pi_j)^{\mathbb{I}(x_1=j)} \prod_{t=2}^{T}\prod_{j=1}^{K}\prod_{k=1}^{K}(A_{jk})^{\mathbb{I}(x_t=k, x_{t-1}=j)}$$

- Hence we get:

$$\hat{\pi}_j = \frac{N_j^1}{\sum_j N_j^1} \qquad \hat{A}_{jk} = \frac{N_{jk}}{\sum_k N_{jk}}$$

where

$$N_j^1 \triangleq \sum_{i=1}^{N}\mathbb{I}(x_{i1} = j), \qquad N_{jk} \triangleq \sum_{i=1}^{N}\sum_{t=1}^{T_i-1}\mathbb{I}(x_{i,t} = j, x_{i,t+1} = k)$$

References I

[Mur12] Kevin P Murphy. *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.

[Mar14] Stephen Marsland. *Machine Learning, An Algorithmic Perspective*. CRC Press, 2014.

[Ian17] Aaron Courville Ian Goodfellow Yoshua Bengio. *Deep Learning*. MIT Press, 2017.