

8 Graphical Models

8.1 Bayesian Networks

Definition. Concepts

1. **Graphical Model** The joint distribution defined by the graph is given by the product, over all nodes of the graph, of a conditional distribution for each node conditioned on the variable corresponding to the parents of that node in the graph

$$p(\mathbf{x}) = \prod_{k=1}^K p(x_k | \text{parent}(x_k))$$

which represents **factorization properties** of the joint distribution for a directed graphical model

(a) Node

i. **Open Circle** Random variable

A. **Shaded** Observed variable, i.e. variable $\{t_n\}$ from the training set, for setting the variable to some observed value

B. **Unshaded** Latent variable

ii. **Solid Circle (Dot)** Deterministic parameter

(b) **Link** probabilistic relationships between these variables

(c) **Plate** labelled with N indicate N nodes of a certain kind by drawing a single representative node and surround it with a box

2. Bayesian Polynomial Regression

$$p(\mathbf{t}, \mathbf{w}) = p(\mathbf{w}) \prod_{n=1}^N p(t_n | \mathbf{w}) \quad p(\mathbf{t}, \mathbf{w} | \mathbf{x}, \alpha, \sigma^2) = p(\mathbf{w} | \alpha) \prod_{n=1}^N p(t_n | \mathbf{w}, x_n, \sigma^2)$$

3. Generative Models

4. Discrete Variable

9 Mixture Models and EM

9.1 K-means Clustering

Definition. Concepts

1. Clustering

2. **K-means** Problem with K-means is that at every step, every point is assigned to one and only one cluster centers, but maybe many points which are halfway between cluster centers. It is beneficial to have soft assignment of such points by adopting a probabilistic viewpoint

3. **Application** in image segmentation and compression

9.2 Mixture of Gaussians

Definition. *Mixtures of Gaussians* Given

$$p(\mathbf{x}) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

Introduce latent variable \mathbf{z}

9.3 An alternative view of EM

Definition. *Concepts*

1. **EM model** Given observed data \mathbf{X} and latent variables \mathbf{Z} , we want to optimize for the likelihood function

$$\ln p(\mathbf{X} | \theta) = \ln \sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z} | \theta)$$