Supplementary Material for Bayesian Network Problem in Project 1

Jun Chu

September 21, 2017

1 Linear Gaussian Model

Consider a Bayesian network where continuous random variable Y has k parents $\mathbf{X} = \{X_1, ..., X_k\}$. Y is said to obey a linear Gaussian model with parameters $\beta_1, ..., \beta_k$ and σ^2 if $P(Y|X_1, ..., X_k) \sim \mathcal{N}\left(\beta_0 + \beta_1 X_1 + ... + \beta_k X_k; \sigma^2\right)$. Let's define $\boldsymbol{\theta} = \left(\beta_1, ..., \beta_k, \sigma^2\right)$. The log-likelihood is Linear Gaussian Model

$$
\begin{aligned}
L(\boldsymbol{\theta}) &= \log\left(P\left(Y|X_1, ..., X_k\right)\right) \\
&= \sum_{n=1}^{N} \left[-\frac{1}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\left(\beta_0 x_0[n] + \beta_1 x_1[n] + ... + \beta_k x_k[n] - y[n]\right)^2\right].
\end{aligned}
$$

Here, we assume $x_0[n] \equiv 1$. It is possible to get a closed-form solution. It involves solving a set of equations which are obtained by taking partial derivatives of the log-likelihood function, as follows. Gradient of log-likelihood with respect to β_i is

$$
\begin{aligned}
\frac{\partial L(\boldsymbol{\theta})}{\partial \beta_i} &= -\frac{1}{\sigma^2}\sum_{n=1}^{N}\left\{\left(\beta_0 x_0[n] + \beta_1 x_1[n] + ... + \beta_k x_k[n] - y[n]\right) x_i[n]\right\} \\
&= -\frac{1}{\sigma^2}\left(\beta_0 \sum_{n=1}^{N} x_0[n]x_i[n] + \beta_1 \sum_{n=1}^{N} x_1[n]x_i[n] + ... + \beta_k \sum_{n=1}^{N} x_k[n]x_i[n] - \sum_{n=1}^{N} y[n]x_i[n]\right).
\end{aligned}
$$

Equating to zero and rearranging we get

$$
\beta_0 \sum_{n=1}^{N} x_0[n]x_i[n] + \beta_1 \sum_{n=1}^{N} x_1[n]x_i[n] + ... + \beta_k \sum_{n=1}^{N} x_k[n]x_i[n] = \sum_{n=1}^{N} y[n]x_i[n], \quad i = 0, 1, ..., k.
$$

If we define:

$$
A = \begin{pmatrix}
\sum_{n=1}^{N} x_0[n]x_0[n] & \sum_{n=1}^{N} x_1[n]x_0[n] & ... & \sum_{n=1}^{N} x_k[n]x_0[n] \\
\sum_{n=1}^{N} x_0[n]x_1[n] & \sum_{n=1}^{N} x_1[n]x_1[n] & ... & \sum_{n=1}^{N} x_k[n]x_1[n] \\
\vdots & \vdots & & \vdots \\
\sum_{n=1}^{N} x_0[n]x_k[n] & \sum_{n=1}^{N} x_1[n]x_k[n] & ... & \sum_{n=1}^{N} x_k[n]x_k[n]
\end{pmatrix},
$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$$

and

$$\boldsymbol{y} = \begin{pmatrix} \sum_{n=1}^{N} y[n]x_0[n] \\ \sum_{n=1}^{N} y[n]x_1[n] \\ \vdots \\ \sum_{n=1}^{N} y[n]x_k[n] \end{pmatrix}.$$

Therefore $A\boldsymbol{\beta} = \boldsymbol{y}$. $\boldsymbol{\beta}$ could be solved using $\boldsymbol{\beta} = A^{-1}\boldsymbol{y}$.

Gradient of log-likelihood with respect to σ is

$$\frac{\partial L(\boldsymbol{\theta})}{\partial \sigma} = -\frac{N}{\sigma} + \frac{1}{\sigma^3} \sum_{n=1}^{N} \left(\beta_0 x_0[n] + \beta_1 x_1[n] + \ldots + \beta_k x_k[n] - y[n]\right)^2.$$

Set it to zero, we get

$$\sigma^2 = \frac{1}{N} \sum_{n=1}^{N} \left(\beta_0 x_0[n] + \beta_1 x_1[n] + \ldots + \beta_k x_k[n] - y[n]\right)^2.$$

2 Bayesian Network Factorization

Given a Bayesian network G of d variables $\mathbf{X} = \{X_1, X_2, ..., X_d\}$, the joint probability distribution is given by

$$p(\mathbf{X}) = \prod_{i=1}^{d} p\left(X_i | \mathrm{pa}(X_i)\right),$$

where $\mathrm{pa}(x_i)$ are the parent variables of x_i. The log-likelihood is

$$\log p(\mathbf{X}) = \sum_{i=1}^{d} \log p\left(X_i | \mathrm{pa}(X_i)\right).$$

Each term could be maximized separately using the aforementioned linear gaussian model closed form solution.