Tutorial 9 - Gaussian Bayesian Networks

COMP9418 – Advanced Topics in Statistical Machine Learning

Lecturer: Gustavo Batista

Lecture: Gaussian Bayesian Networks

Topic: Questions from lecture topics

Last revision: Sunday 7th November, 2021 at 16:22

Recall the equation for a 1 dimensional Gaussian probability distribution function.

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Also recall that the integral of this function is equal to 1.

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = 1$$

Question 1

Let's solve a small inference problem with Gaussian variables. Imagine you are traveling along down a long straight road, and want to know your location. Your location can be described by a distance between 0km and 1000km. We will call your location X. You currently think you are about 30km down the road, but you're quite uncertain. In fact, your belief about your current location can be represented with the distribution $P(X) = \mathcal{N}(30, 4^2)$. You receive a GPS signal that says your current location is 36km. However, you know that GPS signals can be quite noisy (in this imaginary universe), and are usually only accurate to about 1km. We can represent the information you got from the GPS signal as a distribution $P(GPS|X = x) = \mathcal{N}(x, 1^2)$. How can we combine these two sources of information into one distribution that represents your new belief about your location?

Note that this situation corresponds to the Bayesian network: $(X) \rightarrow (GPS)$. You are making a query about X, given the evidence that $GPS = 36$.

Hint: We can use Bayes Theorem to solve this problem. Also, when doing algebraic manipulations with Gaussian distributions, we can forget about the normalizing constant (the fraction at the start of the equation that doesn't contain x) while we do the calculations, then once we have the equation back into a familiar form, we can simply work out what the constant is and put it back at the start. This saves a lot of algebra.

Answer

$$
\begin{aligned}
P(X|GPS=36) &= \frac{P(\text{GPS}=36|X)P(X)}{\int P(\text{GPS}=36|X)P(X)dX} \\
&\propto P(\text{GPS}=36|X)P(X) \\
&\propto \exp\left(-\frac{(36-x)^2}{2\times 1^2}\right)\exp\left(-\frac{(x-30)^2}{2\times 4^2}\right) \\
&= \exp\left(-\frac{(36-x)^2}{2\times 1^2}\right)\exp\left(-\frac{(x-30)^2}{2\times 4^2}\right) \\
&= \exp\left(-\frac{36^2-2\times 36x+x^2}{2\times 1^2}\right)\exp\left(-\frac{x^2-2\times 30x+30^2}{2\times 4^2}\right) \\
&= \exp\left(-(\frac{36^2-2\times 36x+x^2}{2}+\frac{x^2-2\times 30x+30^2}{32})\right) \\
&= \exp\left(-\frac{32(36^2-2\times 36x+x^2)+2(x^2-2\times 30x+30^2)}{32\times 2}\right) \\
&= \exp\left(-\frac{32\times 36^2-32\times 2\times 36x+32x^2+2x^2-2\times 2\times 30x+2\times 30^2}{32\times 2}\right) \\
&= \exp\left(-\frac{(32+2)x^2-(32\times 2\times 36+2\times 2\times 30)x+(32\times 36^2+2\times 30^2)}{32\times 2}\right) \\
&= \exp\left(-\frac{34x^2-2424x+43272}{64}\right) \\
&= \exp\left(-\frac{x^2-(2424/34)x+(43272/34)}{64/34}\right) \\
&\propto \exp\left(-\frac{x^2-2\times(1212/34)x}{64/34}\right) \\
&\propto \exp\left(-\frac{x^2-2\times(1212/34)x+(1212/34)^2}{64/34}\right) \\
&= \exp\left(-\frac{(x-(1212/34))^2}{64/34}\right) \\
&= \exp\left(-\frac{(x-35.647)^2}{2\times 0.97^2}\right) \\
&\propto \frac{1}{0.97\sqrt{2\pi}}\exp\left(-\frac{(x-35.647)^2}{2\times 0.97^2}\right) \\
&= \mathcal{N}(35.647, 0.97^2)
\end{aligned}
$$

This solution shows how we can combine two uncertain sources of information about your location into one estimate, and become more certain about the accuracy of your new estimate than you would be if you relied on either of the other estimates by themselves.

Question 2

The PDF of a multivariate Gaussian distribution is usually written as:

$$p(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}}\exp\left(-\frac{(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{2}\right)$$

where n is the dimension of the vector $\boldsymbol{x}$, and det returns the determinant of a matrix. We will be rearranging this equation into a different form, which we call the canonical form. This form allows us to represent

Gaussian distributions, Gaussian linear conditional distributions, and makes some of the computations easier:

$$p(\boldsymbol{x}) = \exp\left(-\frac{1}{2}\boldsymbol{x}^T K \boldsymbol{x} + \boldsymbol{h}^T \boldsymbol{x} + g\right).$$

Using these two equations, work out what K, $\boldsymbol{h}$ and g are in terms of μ and Σ.

Answer

$$\begin{aligned}
p(\boldsymbol{x}) &= \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\left(-\frac{\boldsymbol{x}^T \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) - \boldsymbol{\mu}^T \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{2}\right) \\
p(\boldsymbol{x}) &= \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\left(-\frac{\boldsymbol{x}^T \Sigma^{-1}\boldsymbol{x} - \boldsymbol{x}^T \Sigma^{-1}\boldsymbol{\mu} - \boldsymbol{\mu}^T \Sigma^{-1}\boldsymbol{x} + \boldsymbol{\mu}^T \Sigma^{-1}\boldsymbol{\mu}}{2}\right) \\
p(\boldsymbol{x}) &= \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\left(-\frac{1}{2}\boldsymbol{x}^T \Sigma^{-1}\boldsymbol{x} + \boldsymbol{\mu}^T \Sigma^{-1}\boldsymbol{x} - \frac{1}{2}\boldsymbol{\mu}^T \Sigma^{-1}\boldsymbol{\mu}\right) \\
p(\boldsymbol{x}) &= \exp\left(-\frac{1}{2}\boldsymbol{x}^T \Sigma^{-1}\boldsymbol{x} + \boldsymbol{\mu}^T \Sigma^{-1}\boldsymbol{x} - \frac{1}{2}\boldsymbol{\mu}^T \Sigma^{-1}\boldsymbol{\mu} - \log\left(\sqrt{(2\pi)^n \det(\Sigma)}\right)\right)
\end{aligned}$$

Therefore $K = \Sigma^{-1}$, $\boldsymbol{h}^T = \mu^T \Sigma^{-1}$ and $g = -\frac{1}{2}\boldsymbol{\mu}^T \Sigma^{-1}\boldsymbol{\mu} - \log\left(\sqrt{(2\pi)^n \det(\Sigma)}\right)$.

Question 3

Let's derive the join operation for two canonical form factors over the same domain. $\phi_3(A, B, C) = \phi_1(A, B, C) \times \phi_2(A, B, C)$. The first factor has parameters $\boldsymbol{K}_1$, $\boldsymbol{h}_1$ and g_1, and the second has $\boldsymbol{K}_2$, $\boldsymbol{h}_2$ and g_2. Work out the parameters of ϕ_3.

Answer

$$\begin{aligned}
\phi_3 &= \exp\left(-\frac{1}{2}\boldsymbol{x}^T K_1 \boldsymbol{x} + \boldsymbol{h}_1^T \boldsymbol{x} + g_1\right) \exp\left(-\frac{1}{2}\boldsymbol{x}^T K_2 \boldsymbol{x} + \boldsymbol{h}_2^T \boldsymbol{x} + g_2\right) \\
&= \exp\left(-\frac{1}{2}\boldsymbol{x}^T (K_1 + K_2)\boldsymbol{x} + (\boldsymbol{h}_1^T + \boldsymbol{h}_2^T)\boldsymbol{x} + (g_1 + g_2)\right)
\end{aligned}$$

Therefore the parameters are $K_3 = K_1 + K_2$, $h_3 = h_1 + h_2$, and $g_3 = g_1 + g_2$.

Question 4

Now we want to be able to multiply factors that have different domains. To do this we need to `extend` our factor by adding variables to the domain of the factor (without changing the meaning of the factor). To see how we do this, prove that:

$$\exp\left(-\frac{1}{2}\begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} h_1 & h_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + g\right)$$

is equal to

$$\exp\left(-\frac{1}{2}\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} K_{11} & K_{12} & 0 \\ K_{21} & K_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} h_1 & h_2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + g\right).$$

Answer Noting that the inverse covariance matrix is symmetric.

$$\begin{aligned}
\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} K_{11} & K_{12} & 0 \\ K_{21} & K_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} &= K_{11}x_1^2 + 2K_{12}x_1x_2 + K_{22}x_2^2 + 0x_1x_3 + 0x_2x_3 + 0x_3^2 \\
&= K_{11}x_1^2 + 2K_{12}x_1x_2 + K_{22}x_2^2 \\
&= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}
\end{aligned}$$

$$\begin{bmatrix} h_1 & h_2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = h_1 x_1 + h_2 x_2 = \begin{bmatrix} h_1 & h_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$