gaussian identities

sam roweis

(revised July 1999)

0.1 multidimensional gaussian

a $d$-dimensional multidimensional gaussian (normal) density for $\mathbf{x}$ is:

$$
\mathcal{N}(\mu, \Sigma) = (2\pi)^{-d/2} |\Sigma|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)\right]
$$

(1)

it has entropy:

$$
S = \frac{1}{2} \log_2 \left((2\pi e)^d |\Sigma|\right) - \text{const bits}
$$

(2)

where $\Sigma$ is a symmetric positive semi-definite covariance matrix and the (unfortunate) constant is the log of the units in which $\mathbf{x}$ is measured over the “natural units”

0.2 linear functions of a normal vector

no matter how $\mathbf{x}$ is distributed,

$$
E[\mathbf{A}\mathbf{x} + \mathbf{y}] = \mathbf{A}(E[\mathbf{x}]) + \mathbf{y}
$$

(3a)

$$
\text{Covar}[\mathbf{A}\mathbf{x} + \mathbf{y}] = \mathbf{A}(\text{Covar}[\mathbf{x}])\mathbf{A}^T
$$

(3b)

in particular this means that for normal distributed quantities:

$$
\mathbf{x} \sim \mathcal{N}(\mu, \Sigma) \Rightarrow (\mathbf{A}\mathbf{x} + \mathbf{y}) \sim \mathcal{N}(\mathbf{A}\mu + \mathbf{y}, \mathbf{A}\Sigma\mathbf{A}^T)
$$

(4a)

$$
\mathbf{x} \sim \mathcal{N}(\mu, \Sigma) \Rightarrow \Sigma^{-1/2}(\mathbf{x} - \mu) \sim \mathcal{N}(\mathbf{0}, \mathbf{I})
$$

(4b)

$$
\mathbf{x} \sim \mathcal{N}(\mu, \Sigma) \Rightarrow (\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu) \sim \chi^2_n
$$

(4c)
0.3 marginal and conditional distributions

Let the vector \( z = [x^T y^T]^T \) be normally distributed according to:

\[
z = \begin{bmatrix} x \\ y \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} a \\ b \end{bmatrix}, \begin{bmatrix} A & C \\ C^T & B \end{bmatrix} \right)
\]

(5a)

where \( C \) is the (non-symmetric) cross-covariance matrix between \( x \) and \( y \) which has as many rows as the size of \( x \) and as many columns as the size of \( y \). Then the marginal distributions are:

\[
x \sim \mathcal{N}(a, A) \quad (5b)
\]
\[
y \sim \mathcal{N}(b, B) \quad (5c)
\]

and the conditional distributions are:

\[
x \vert y \sim \mathcal{N}(a + CB^{-1}(y - b), A - CB^{-1}C^T) \quad (5d)
\]
\[
y \vert x \sim \mathcal{N}(b + C^T A^{-1}(x - a), B - C^T A^{-1}C) \quad (5e)
\]

0.4 multiplication

The multiplication of two gaussian functions is another gaussian function (although no longer normalized). In particular,

\[
\mathcal{N}(a, A) \cdot \mathcal{N}(b, B) \propto \mathcal{N}(c, C)
\]

(6a)

where

\[
C = (A^{-1} + B^{-1})^{-1} \quad (6b)
\]
\[
c = CA^{-1}a + CB^{-1}b \quad (6c)
\]

Amazingly, the normalization constant \( z_c \) is Gaussian in either \( a \) or \( b \):

\[
z_c = (2\pi)^{-d/2} |C|^{1/2} |A|^{-1/2} |B|^{-1/2} \exp \left[ -\frac{1}{2}(a^T A^{-1}a + b^T B^{-1}b - c^T C^{-1}c) \right]
\]

(6d)

\[
z_c(a) \sim \mathcal{N}( (A^{-1}CA^{-1})^{-1}(A^{-1}CB^{-1})b, (A^{-1}CA^{-1})^{-1} ) \quad (6e)
\]
\[
z_c(b) \sim \mathcal{N}( (B^{-1}CB^{-1})^{-1}(B^{-1}CA^{-1})a, (B^{-1}CB^{-1})^{-1} ) \quad (6f)
\]
0.5 quadratic forms

the expectation of a quadratic form under a gaussian is another quadratic form (plus an annoying constant). in particular, if \( x \) is gaussian distributed with mean \( m \) and variance \( S \) then,

\[
\int \frac{(x - \mu)^T \Sigma^{-1}(x - \mu)}{x} \mathcal{N}(m, S) \, dx
= (\mu - m)^T \Sigma^{-1}(\mu - m) + \text{Tr} \left[ \Sigma^{-1} S \right]
\]  

(7a)

if the original quadratic form has a linear function of \( x \) the result is still simple:

\[
\int \frac{(Wx - \mu)^T \Sigma^{-1}(Wx - \mu)}{x} \mathcal{N}(m, S) \, dx
= (\mu - Wm)^T \Sigma^{-1}(\mu - Wm) + \text{Tr} \left[ W^T \Sigma^{-1} WS \right]
\]  

(7b)

0.6 convolution

the convolution of two gaussian functions is another gaussian function (although no longer normalized). in particular,

\[
\mathcal{N}(a, A) \ast \mathcal{N}(b, B) \propto \mathcal{N}(a + b, A + B)
\]  

(8)

this is a direct consequence of the fact that the Fourier transform of a gaussian is another gaussian and that the multiplication of two gaussians is still gaussian.

0.7 Fourier transform

the (inverse)Fourier transform of a gaussian function is another gaussian function (although no longer normalized). in particular,

\[
\mathcal{F}[\mathcal{N}(a, A)] \propto \mathcal{N}(jA^{-1}a, A^{-1})  
\]  

(9a)

\[
\mathcal{F}^{-1}[\mathcal{N}(b, B)] \propto \mathcal{N}(-jB^{-1}b, B^{-1})  
\]  

(9b)

where \( j = \sqrt{-1} \)
0.8 constrained maximization

the maximum over $x$ of the quadratic form:

$$\mu^T x - \frac{1}{2} x^T A^{-1} x$$

subject to the $J$ conditions $c_j(x) = 0$ is given by:

$$A\mu + ACA, \quad \Lambda = -4(C^T AC)C^T A\mu$$

where the $j$th column of $C$ is $\partial c_j(x)/\partial x$