gaussian identities

sam roweis

(revised July 1999)

0.1 multidimensional gaussian

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

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

it has entropy:

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

where $\boldsymbol{\Sigma}$ is a symmetric postive 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,

$$\mathrm{E}[\mathbf{Ax}+\mathbf{y}] = \mathbf{A}(\mathrm{E}[\mathbf{x}]) + \mathbf{y} \tag{3a}$$

$$\mathrm{Covar}[\mathbf{Ax}+\mathbf{y}] = \mathbf{A}(\mathrm{Covar}[\mathbf{x}])\mathbf{A}^T \tag{3b}$$

in particular this means that for normal distributed quantities:

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Rightarrow (\mathbf{Ax}+\mathbf{y}) \sim \mathcal{N}\left(\mathbf{A}\boldsymbol{\mu}+\mathbf{y}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T\right) \tag{4a}$$

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Rightarrow \boldsymbol{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu}) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \tag{4b}$$

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Rightarrow (\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \sim \chi^2_n \tag{4c}$$

0.3 marginal and conditional distributions

let the vector $\mathbf{z} = [\mathbf{x}^T \mathbf{y}^T]^T$ be normally distributed according to:

$$\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{C}^T & \mathbf{B} \end{bmatrix} \right) \tag{5a}$$

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

$$\mathbf{x} \sim \mathcal{N}(\mathbf{a}, \mathbf{A}) \tag{5b}$$

$$\mathbf{y} \sim \mathcal{N}(\mathbf{b}, \mathbf{B}) \tag{5c}$$

and the conditional distributions are:

$$\mathbf{x}|\mathbf{y} \sim \mathcal{N}\left(\mathbf{a} + \mathbf{C}\mathbf{B}^{-1}(\mathbf{y} - \mathbf{b}), \mathbf{A} - \mathbf{C}\mathbf{B}^{-1}\mathbf{C}^T\right) \tag{5d}$$

$$\mathbf{y}|\mathbf{x} \sim \mathcal{N}\left(\mathbf{b} + \mathbf{C}^T\mathbf{A}^{-1}(\mathbf{x} - \mathbf{a}), \mathbf{B} - \mathbf{C}^T\mathbf{A}^{-1}\mathbf{C}\right) \tag{5e}$$

0.4 multiplication

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

$$\mathcal{N}(\mathbf{a}, \mathbf{A}) \cdot \mathcal{N}(\mathbf{b}, \mathbf{B}) \propto \mathcal{N}(\mathbf{c}, \mathbf{C}) \tag{6a}$$

where

$$\mathbf{C} = \left(\mathbf{A}^{-1} + \mathbf{B}^{-1}\right)^{-1} \tag{6b}$$

$$\mathbf{c} = \mathbf{C}\mathbf{A}^{-1}\mathbf{a} + \mathbf{C}\mathbf{B}^{-1}\mathbf{b} \tag{6c}$$

amazingly, the normalization constant z_c is Gaussian in either $\mathbf{a}$ or $\mathbf{b}$:

$$z_c = (2\pi)^{-d/2}|\mathbf{C}|^{+1/2}|\mathbf{A}|^{-1/2}|\mathbf{B}|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{a}^T\mathbf{A}^{-1}\mathbf{a} + \mathbf{b}^T\mathbf{B}^{-1}\mathbf{b} - \mathbf{c}^T\mathbf{C}^{-1}\mathbf{c})\right] \tag{6d}$$

$$z_c(\mathbf{a}) \sim \mathcal{N}\left((\mathbf{A}^{-1}\mathbf{C}\mathbf{A}^{-1})^{-1}(\mathbf{A}^{-1}\mathbf{C}\mathbf{B}^{-1})\mathbf{b}, (\mathbf{A}^{-1}\mathbf{C}\mathbf{A}^{-1})^{-1}\right) \tag{6e}$$

$$z_c(\mathbf{b}) \sim \mathcal{N}\left((\mathbf{B}^{-1}\mathbf{C}\mathbf{B}^{-1})^{-1}(\mathbf{B}^{-1}\mathbf{C}\mathbf{A}^{-1})\mathbf{a}, (\mathbf{B}^{-1}\mathbf{C}\mathbf{B}^{-1})^{-1}\right) \tag{6f}$$

0.5 quadratic forms

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

$$\int_{\mathbf{x}} (\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu}) \mathcal{N}(\mathbf{m},\mathbf{S})\, d\mathbf{x} = (\boldsymbol{\mu}-\mathbf{m})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}-\mathbf{m}) + \mathrm{Tr}\left[\boldsymbol{\Sigma}^{-1}\mathbf{S}\right] \tag{7a}$$

if the original quadratic form has a linear function of $\mathbf{x}$ the result is still simple:

$$\int_{\mathbf{x}} (\mathbf{W}\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{W}\mathbf{x}-\boldsymbol{\mu}) \mathcal{N}(\mathbf{m},\mathbf{S})\, d\mathbf{x} = (\boldsymbol{\mu}-\mathbf{W}\mathbf{m})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}-\mathbf{W}\mathbf{m}) + \mathrm{Tr}\left[\mathbf{W}^T\boldsymbol{\Sigma}^{-1}\mathbf{W}\mathbf{S}\right] \tag{7b}$$

0.6 convolution

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

$$\mathcal{N}(\mathbf{a},\mathbf{A}) * \mathcal{N}(\mathbf{b},\mathbf{B}) \propto \mathcal{N}(\mathbf{a}+\mathbf{b},\mathbf{A}+\mathbf{B}) \tag{8}$$

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,

$$\mathcal{F}\left[\mathcal{N}(\mathbf{a},\mathbf{A})\right] \propto \mathcal{N}\left(j\mathbf{A}^{-1}\mathbf{a},\mathbf{A}^{-1}\right) \tag{9a}$$

$$\mathcal{F}^{-1}\left[\mathcal{N}(\mathbf{b},\mathbf{B})\right] \propto \mathcal{N}\left(-j\mathbf{B}^{-1}\mathbf{b},\mathbf{B}^{-1}\right) \tag{9b}$$

where $j=\sqrt{-1}$

0.8 constrained maximization

the maximum over $\mathbf{x}$ of the quadratic form:

$$\boldsymbol{\mu}^T \mathbf{x} - \frac{1}{2} \mathbf{x}^T \mathbf{A}^{-1} \mathbf{x} \tag{10a}$$

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

$$\mathbf{A}\boldsymbol{\mu} + \mathbf{A}\mathbf{C}\boldsymbol{\Lambda}, \qquad \boldsymbol{\Lambda} = -4(\mathbf{C}^T \mathbf{A}\mathbf{C})\mathbf{C}^T \mathbf{A}\boldsymbol{\mu} \tag{10b}$$

where the jth column of $\mathbf{C}$ is $\partial c_j(\mathbf{x})/\partial \mathbf{x}$