5SSD0 formula sheet

6 February 2025

Distributions

Gaussian distribution

The (*moment* parameterization of the) **Gaussian distribution** is given by

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

Alternatively, the *canonical* parameterization of the Gaussian distribution is given by

$$\mathcal{N}_c(x|\eta,\Lambda) = \exp\left(a + \eta^T x - \frac{1}{2} x^T \Lambda x \right) ,$$

where where $\eta = \Sigma^{-1}\mu$ is the canonical mean, $\Lambda = \Sigma^{-1}$ is the precision matrix, and a is a normalization constant.

Multiplication of Gaussians

$$\mathcal{N}(x|\mu_a,\Sigma_a)\mathcal{N}(x|\mu_b,\Sigma_b) = \underbrace{\mathcal{N}(\mu_a|\mu_b,\Sigma_a+\Sigma_b)}_{\text{normalization constant}}\mathcal{N}(x|\mu_c,\Sigma_c)$$

where $\Sigma_c^{-1} = \Sigma_a^{-1} + \Sigma_b^{-1}$, and $\Sigma_c^{-1}\mu_c = \Sigma_a^{-1}\mu_a + \Sigma_b^{-1}\mu_b$, i.e., precisions add and precision-weighted means add when multiplying two Gaussian distributions.

Beta distribution

For a variable $x \in [0,1]$, the **Beta distribution** is given by

$$\text{beta}(x|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} .$$

The expected value of the beta distribution is $\mathrm{E}\left[\text{beta}(x|\alpha,\beta)\right] = \frac{\alpha}{\alpha+\beta}$.

Matrix Calculus

We define the **gradient** of a scalar function $f(A)$ with respect to an $n \times k$ matrix A as

$$\nabla_A f = \begin{bmatrix} \frac{\partial f}{\partial a_{11}} & \frac{\partial f}{\partial a_{12}} & \cdots & \frac{\partial f}{\partial a_{1k}} \\ \frac{\partial f}{\partial a_{21}} & \frac{\partial f}{\partial a_{22}} & \cdots & \frac{\partial f}{\partial a_{2k}} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial f}{\partial a_{n1}} & \frac{\partial f}{\partial a_{n2}} & \cdots & \frac{\partial f}{\partial a_{nk}} \end{bmatrix}$$

The following maxtrix calculus formulas are useful:

$$\begin{aligned} |A^{-1}| &= |A|^{-1} \\ \nabla_A \log |A| &= (A^T)^{-1} = (A^{-1})^T \\ \mathrm{Tr}[ABC] &= \mathrm{Tr}[CAB] = \mathrm{Tr}[BCA] \\ \nabla_A \mathrm{Tr}[AB] &= \nabla_A \mathrm{Tr}[BA] = B^T \\ \nabla_A \mathrm{Tr}[ABA^T] &= A(B + B^T) \\ \nabla_x x^T A x &= (A + A^T)x \\ \nabla_X a^T X b &= \nabla_X \mathrm{Tr}[ba^T X] = ab^T \end{aligned}$$