Multivariate Gaussian

$$p(x|\mu,\Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\}$$

- Moment Parameterization: $\mu = \mathbb{E}(X)$, $\Sigma = \texttt{Cov}(X) = \mathbb{E}[(X-\mu)(X-\mu)^T]$ (symmetric, positive semi-definite matrix).
- Mahalanobis distance: $\triangle^2 = (x-\mu)^T\Sigma^{-1}(x-\mu)$.
- Canonical Parameterization:

$$p(x|\eta,\Lambda) = \exp\{a + \eta^T x - \frac{1}{2}x^T\Lambda x\}$$

where $\Lambda = \Sigma^{-1}$, $\eta = \Sigma^{-1}\mu$, $a = -\frac{1}{2}\left(n\log 2\pi - \log|\Lambda| + \eta^T\Lambda\eta\right)$.
- Tons of applications (MoG, FA, PPCA, Kalman Filter, ...)

Multivariate Gaussian $P(X_1, X_2)$

$P(X_1, X_2)$ (Joint Gaussian)

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

$P(X_2)$ (Marginal Gaussian)

$$\mu_2^m = \mu_2, \quad \Sigma_2^m = \Sigma_2$$

$P(X_1 | X_2 = x_2)$ (Conditional Gaussian)

$$\begin{aligned} \mu_{1|2} &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) \\ \Sigma_{1|2} &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \end{aligned}$$

Operations on Gaussian R.V.

The linear transform of a gaussian r.v. is a guassian. Remember that no matter how x is distributed,

$$\mathbb{E}(AX + b) = A\mathbb{E}(X) + b$$

$$\texttt{Cov}(AX + b) = A\texttt{Cov}(X)A^T$$

this means that for gaussian distributed quantities:

$$X \sim \mathcal{N}(\mu, \Sigma) \quad \Rightarrow \quad AX + b \sim \mathcal{N}(A\mu + b, A\Sigma A^T).$$

The sum of two independent gaussian r.v. is a gaussian.

$$Y = X_1 + X_2, X_1 \perp X_2 \quad \Rightarrow \quad \mu_Y = \mu_1 + \mu_2, \quad \Sigma_Y = \Sigma_1 + \Sigma_2$$

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

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

$$\text{where } C = (A^{-1} + B^{-1})^{-1}, c = CA^{-1}a + CB^{-1}b$$

Maximum Likelihood Estimate of μ and Σ

Given a set of i.i.d. data $X = \{x_1, \ldots, x_N\}$ drawn from $\mathcal{N}(x; \mu, \Sigma)$, we want to estimate (μ, Σ) by MLE. The log-likelihood function is

$$\ln p(X|\mu, \Sigma) = -\frac{N}{2} \ln |\Sigma| - \frac{1}{2} \sum_{n=1}^{N} (x_n - \mu)^T \Sigma^{-1} (x_n - \mu) + \text{const}$$

Taking its derivative w.r.t. μ and setting it to zero we have

$$\hat{\mu} = \frac{1}{N} \sum_{n=1}^{N} x_n$$

Rewrite the log-likelihood using "trace trick",

$$\begin{aligned} \ln p(X|\mu, \Sigma) &= -\tfrac{N}{2} \ln |\Sigma| - \tfrac{1}{2} \sum_{n=1}^{N} (x_n - \mu)^T \Sigma^{-1} (x_n - \mu) + \text{const} \\ &\propto -\tfrac{N}{2} \ln |\Sigma| - \tfrac{1}{2} \sum_{n=1}^{N} \text{Trace} \left(\Sigma^{-1} (x_n - \mu)(x_n - \mu)^T \right) \\ &= -\tfrac{N}{2} \ln |\Sigma| - \tfrac{1}{2} \text{Trace} \left(\Sigma^{-1} \sum_{n=1}^{N} [(x_n - \mu)(x_n - \mu)^T] \right) \end{aligned}$$

Taking the derivative w.r.t. Σ^{-1}, and using 1) $\frac{\partial}{\partial A} \log |A| = A^{-T}$; 2) $\frac{\partial}{\partial A} \mathtt{Tr}[AB] = \frac{\partial}{\partial A} \mathtt{Tr}[BA] = B^T$, we obtain

$$\hat{\Sigma} = \frac{1}{N} \sum_{n=1}^{N} (x_n - \hat{\mu}) (x_n - \hat{\mu})^T .$$