Lecture Notes Summary

Complete derivation of the 0-Entropic prior for the k-dim Gaussians.

Topics Covered

The Manifold of k-dim Gaussians I

Let $x, \mu \in R^k$ and $\Sigma \in R^{k\times k}$ symmetric and positive definite. We write $\Sigma > 0$ to indicate that Σ is symmetric and positive definite. We say that $X \sim \mathrm{N}(\mu, \Sigma)$, i.e. it is $k-$dim gaussian with mean vector μ and variance matrix Σ when the pdf (w.r.t. Lebesgue measure dx in R^k) is,

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

Information Separation

Let $\theta_0 = (\mu_0, \Sigma_0)$ and $\theta = (\mu, \Sigma)$. Then the Kullback number between the $P_0 = N(\mu_0, \Sigma_0)$ and $P = N(\mu, \Sigma)$ is denoted by $\mathrm{I}(\theta_0 : \theta)$. By defininiton,

$$\mathrm{I}(\theta_0 : \theta) = \mathrm{E}_0\left(\log\frac{P_0(dx)}{P(dx)}\right)$$

and for the k-gaussians we have,

$$\mathrm{I}(\theta_0 : \theta) = \frac{1}{2}\log|\Sigma_0^{-1}\Sigma| + \frac{1}{2}\mathrm{E}_0\left((x-\mu)^{'}\Sigma^{-1}(x-\mu) - (x-\mu_0)^{'}\Sigma_0^{-1}(x-\mu_0)\right)$$

The Trace Trick:

The trace of a scalar is just the scalar itself and $\mathrm{tr}(AB) = \mathrm{tr}(BA)$. This implies the very useful "trace trick":

$$\mathrm{E}_0\left((x-\mu_0)^{'}\Sigma_0^{-1}(x-\mu_0)\right) = \mathrm{E}_0\ \mathrm{tr}\left((x-\mu_0)(x-\mu_0)^{'}\Sigma_0^{-1}\right) = \mathrm{tr}\left(\Sigma_0\Sigma_0^{-1}\right) = k$$

and by adding and substracting μ_0 and using the trace trick again,

$$\mathrm{E}_0\left((x-\mu)^{'}\Sigma^{-1}(x-\mu)\right) = (\mu-\mu_0)^{'}\Sigma^{-1}(\mu-\mu_0) + \mathrm{tr}\left(\Sigma_0\Sigma^{-1}\right).$$

Thus,

$$\mathrm{I}(\theta_0 : \theta) = \frac{1}{2}(\mu-\mu_0)^{'}\Sigma^{-1}(\mu-\mu_0) + \frac{1}{2}\log|\Sigma\Sigma_0^{-1}| + \frac{1}{2}\mathrm{tr}\left(\Sigma_0\Sigma^{-1}\right) - \frac{k}{2}$$

Notice that the dimension of the manifold of k-dim gaussians is the number of independent parameters in $\theta = (\mu, \Sigma)$ i.e. $k + k(k+1)/2$, since $\mu \in R^k$ and σ is symmetric. When $k = 1$ the manifold is of dimension two, and when $k = 2$ it is of dimension 5. Remember the 5 number summary for simple linear regression: The two means, the two SDs and the correlation coefficient.

The Information Metric II

The fastest way to get all the entries of the Fisher information matrix $G(\theta) = (g_{ij}(\theta))$, (i.e. the information metric) is by expanding $J(t) = \mathrm{I}(\theta : \theta + tv)$ in a taylor series about $t = 0$ and using the fact that,

$$J(t) = \frac{t^2}{2} v^{'} G(\theta) v + o(t^2)$$

and thus,

$$J^{''}(0) = v^{'} G(\theta) v$$

gives all the entries g_{ij} at once by simply collecting the coefficients of the terms of the taylor series that are quadratic in t. Clearly only the first term of $\mathrm{I}(\theta_0 : \theta)$ contains μ and it is already quadratic in μ. The other terms contain Σ but not μ and therefore the information matrix must be block diagonal,

$$G(\theta) = \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & Q \end{bmatrix}$$

where $Q > 0$ is of dimension $k(k+1)/2$. Let us separate the components of the velocity vector as (v, V) as in $\theta = (\mu, \Sigma)$. We have,

$$J(t) = \frac{t^2}{2} v^{'} \Sigma^{-1} v + \frac{1}{2} \log \frac{|\Sigma + tV|}{|\Sigma|} + \frac{1}{2} \mathrm{tr}\left(\Sigma (\Sigma + tV)^{-1} \right) - \frac{k}{2}$$

that simplifies to,

$$J(t) = \frac{t^2}{2} v^{'} \Sigma^{-1} v + \frac{1}{2} \log |I + t\Sigma^{-1} V| + \frac{1}{2} \mathrm{tr}\left((I + t\Sigma^{-1} V)^{-1} \right) - \frac{k}{2}$$

We notice that for any square matrix A, we have,

$$|I + tA| = 1 + \mathrm{tr}(A)t + \mathrm{tr}_2(A)t^2 + o(t^2)$$

where tr_2 is the familiar second trace from the standard expansion of the characteristic polynomial of a square matrix. Recall that the second trace is the sum of all the 2 by 2 determinants obtained from A by discarding all but 2 rows and 2 cols of A. It is not difficult to check this formula by induction on the dimension of A. It is also straight forward (again by using induction on the size of A) to check that,

$$\mathrm{tr}_2(A) = \frac{1}{2} \left(\mathrm{tr}^2(A) - \mathrm{tr}(A^2) \right).$$

It is also clear that,

$$(I + tA)^{-1} = I - tA + t^2 A^2 + o(t^2)$$

and hence,

$$\frac{1}{2} \mathrm{tr}(I + tV)^{-1} = \frac{k}{2} - \frac{t}{2} \mathrm{tr}(A) + \frac{t^2}{2} \mathrm{tr}(A^2) + o(t^2)$$

and from the previous expression for the second trace we obtain,

$$\begin{aligned} \frac{1}{2} \log |I + tV| &= \frac{1}{2} \mathrm{tr}(A)t + \frac{1}{2} (2\mathrm{tr}_2(A) - \mathrm{tr}^2(A)) \frac{t^2}{2} + o(t^2) \\ &= \frac{1}{2} \mathrm{tr}(A) - \frac{1}{2} \mathrm{tr}(A^2) \frac{t^2}{2} + o(t^2). \end{aligned}$$

Collecting all the coefficients of t^2 we finally obtain,

$$J^{''}(0) = \frac{t^2}{2} \left(v^{'} \Sigma^{-1} v + \frac{1}{2} \mathrm{tr}(A^2) \right) + o(t^2)$$

where $A = V\Sigma^{-1}$ or $A = \Sigma^{-1}V$ both give the same answer. This allows to compute the entries of $Q = (q_{ij})$. Let us write the lower triangular elements of V in a vector v (do not confuse with the previous v that it won't be needed any more) of dimension $k(k+1)/2$,

$$v^{'} = (v_{11}, v_{21}, v_{22}, v_{31}, v_{32}, v_{33}, \ldots, v_{kk})$$

so that,

$$v^{'}Qv = \frac{1}{2}\text{tr}(A^2) = \frac{1}{2}\text{tr}(V\Sigma^{-1}V\Sigma^{-1}).$$

Now, denote $\Sigma^{-1} = (\sigma^{ij})$ and by e_j the j-th cannonical vector (1 in position j and 0 everywhere else) and by E_{ij} the matrix with all zeroes except for a 1 in the i-th row j-th column. We have,

$$\begin{aligned} q_{ij} = e_i^{'}Qe_j &= \frac{1}{2}\text{tr}(V_i\Sigma^{-1}V_j\Sigma^{-1}) \\ &= \frac{1}{2}\text{tr}\left([E_{i_1i_2} + (1-\delta_{i_1i_2})E_{i_1i_2}]\Sigma^{-1}[E_{j_1j_2} + (1-\delta_{j_1j_2})E_{j_2j_1}]\Sigma^{-1}\right) \end{aligned}$$

where we have used the fact that when the i-th entry in v is $v_{i_1i_2}$ then the corresponding V is,

$$V_i = E_{i_1i_2} + (1-\delta_{i_1i_2})E_{i_1i_2}.$$

expanding the products we obtain,

$$\begin{aligned} q_{ij} = \frac{1}{2}[\sigma^{i_1j_1}\sigma^{i_2j_2} &+ (1-\delta_{j_1j_2})\sigma^{i_1j_2}\sigma^{i_2j_1} \\ &+ (1-\delta_{i_1i_2})\sigma^{i_2j_1}\sigma^{i_1j_2} \\ &+ (1-\delta_{i_1i_2})(1-\delta_{j_1j_2})\sigma^{i_1j_1}\sigma^{i_2j_2}] \end{aligned}$$

Example 1

When $k = 2$. we have,

$$Q = \begin{bmatrix} \frac{1}{2}(\sigma^{11})^2 & \sigma^{11}\sigma^{21} & (\sigma^{12})^2/2 \\ \sigma^{11}\sigma^{21} & \sigma^{22}\sigma^{11} + (\sigma^{11})^2 & \sigma^{22}\sigma^{21} \\ (\sigma^{12})^2/2 & \sigma^{22}\sigma^{21} & \frac{1}{2}(\sigma^{11})^2 \end{bmatrix}$$

In this case we can check directly that,

$$\det Q = |\Sigma^{-1}|^3/4$$

#1

In general, it is possible to show that

$$\det Q = 2^{-k}|\Sigma^{-1}|^{k+1}$$

Proof: A tedious but straight forward strategy may be as follows: Using the explicit formulas for q_{ij}, just prove it first for the case when Σ^{-1} is an elementary matrix, i.e., one of the three types obtainable by performing one of the three elementary row operations to the identity matrix. Recall that the elementary row operations are: swapping two rows, multiplying a row by a non-zero number, and, adding a multiple of one row to another row. Then, one needs to check that $Q(EA) = Q(E)Q(A)$ and finally use the standard fact that every non-singular matrix is the product of elementary matrices.

The Entropic Priors

The 0-Entropic Prior for a regular model M is given by a family of scalar probability density functions on M. For an initial prior guess P_0 and prior number of observations $\alpha > 0$, sufficiently large so that the prior is proper, the scalar density is:

$$\pi(P|P_0,\alpha) = \frac{1}{Z}\exp(-\alpha \mathrm{I}(P_0 : P))$$

where α needs to be large enough so that,

$$Z = \int_M e^{-\alpha\ I(P_0:P)}\, dM \ < \ \infty$$

where the integration is with respect to the volume form in the Riemannian manifold (M, g) with g the information metric. i.e., in a given θ-parametrization, $M = \{P_\theta : \theta \in \Theta\}$,

$$dM = \sqrt{\det g(\theta)}\ d\theta$$

and, the density becomes a function of $\theta \in \Theta$,

$$\pi(\theta|\theta_0,\alpha) = \frac{1}{Z} e^{-\alpha\, I(\theta_0:\theta)}$$

Notice that the choice of $P_0 = P_{\theta_0}$ is just a simple special case.

The 1-Entropic Priors are just like the above but with $\mathrm{I}(\theta : \theta_0)$ in the exponent. It is possible to show that the 0-Entropy, $\mathrm{I}_0 = \mathrm{I}(P_0 : P)$, and the 1-Entropy, $\mathrm{I}_1 = \mathrm{I}(P : P_0)$ are two extremes of a continium of entropies I_δ for $0 < \delta < 1$ that produce δ-Entropic Priors with tails that are not exponential but polynomial that are multivariate generalizations of the student-t distributions.

Entropic Priors for the Multivariate Gaussians

When the model M is the special case of the manifold of k-dimensional Gaussian distributions, the 0-Entropic prior is obtained by applying the general formulas for entropic priors to the specific formulas for the entropy, information metric, and, volume form previously obtained for the k-Gaussians. The scalar density for the 0-Entropic prior is,

$$\begin{aligned}\pi(\mu,\Sigma|\mu_0,\Sigma_0,\alpha) = \frac{1}{Z}&\cdot \exp(-\frac{1}{2}(\mu-\mu_0)^{'}(\frac{1}{\alpha}\Sigma)^{-1}(\mu-\mu_0)) \\ &\cdot |\Sigma^{-1}|^{\alpha/2}\exp(-\frac{\alpha}{2}\mathrm{tr}(\Sigma_0\Sigma^{-1}))\end{aligned}$$

where the normalizing constant is obtained by integrating the above with respect to the Riemannian volume element,

$$dM = \frac{d\mu \wedge d\Sigma}{2^{k/2}|\Sigma|^{(k+2)/2}}$$

and after integrating over $\mu \in R^k$ we have,

$$Z = (\frac{\pi}{\alpha})^{k/2}\int_{\Sigma>0} |\Sigma^{-1}|^{\alpha/2}\exp(-\frac{\alpha}{2}\mathrm{tr}(\Sigma_0\Sigma^{-1}))\,\frac{d\Sigma}{|\Sigma|^{(k+1)/2}}$$

Example 1

Vestibulum tortor.sto.	#1

Example 2

Vestibulum tortor.sto. #2

TopicName

Aliquam erat volutpat. ctum non, mattis id, urna. Suspendisse sed ligula.

Vestibulum eger justo.	#1