Covariance Cleaning

E-mail:

GP's

The Marginal log-likelihood of a GP, for a given covariance matrix $C(\theta) = K(\theta) + \sigma^2 I \in \mathbb{R}^{N\times N}$

$$\mathcal{L} = -\log p(y|\theta) = \frac{1}{2}\Big(\log \det C(\theta) + \langle y|C(\theta)|y\rangle + n\log(2\pi)\Big) \tag{1}$$

Now by the spectral theorem we can write

$$C = \sum_i^N \lambda_i |\phi_i\rangle\langle\phi_i| = \sum_i^m \lambda_i |\phi_i\rangle\langle\phi_i| + \sum_{i=m+1}^N \lambda_i |\phi_i\rangle\langle\phi_i| \tag{2}$$

This gives us two procedures by which we can clean the covariance matrices of GP's and optimise over using iterative methods.

Deleting Eigenvalues

$$C \approx \frac{n}{m}\sum_i^m \lambda_i |\phi_i\rangle\langle\phi_i| \tag{3}$$

and hence

$$\mathcal{L} = -\log p(y|\theta) = \frac{1}{2}\Big(m\log\frac{n}{m} + \sum_i^m \log\lambda_i + \sum_i^m \frac{m}{n\lambda_i}|\langle y|\phi_i\rangle|^2 + m\log(2\pi)\Big) \tag{4}$$

Flooring eigenvalues

$$C \approx \sum_i^m \lambda_i |\phi_i\rangle\langle\phi_i| + \sum_{i=m+1}^n \lambda_0 |z\rangle\langle z| = \sum_i^m \lambda_i |\phi_i\rangle\langle\phi_i| + (n-m)\lambda_0 |z\rangle\langle z| \tag{5}$$

where we make λ_0 so as to keep the trace of the matrix invariant. I.e

$$\lambda_0 = \frac{Tr(C) - \sum_i^m \lambda_i}{n - m} \tag{6}$$

and z is just a random vector. Hence we now have

$$\mathcal{L} = -\log p(y|\theta) = \frac{1}{2}\Bigg((n-m)\log\lambda_0 + \sum_i^m \log\lambda_i + \sum_i^m \frac{1}{\lambda_i}|\langle y|\phi_i\rangle|^2 + \sum_{i=m+1}^{n} \frac{1}{\lambda_0}|\langle y|z\rangle|^2 + n\log(2\pi)\Bigg) \tag{7}$$

Question for Stefan

O.k so there is something about this spectral decomposition notation which is messing with my head, so we know trivially that

$$Tr(C^m) = Tr(IC^m) = Tr([U^tU]^m C^m) = Tr([U^tCU]^m) = Tr(D^m) = \sum_i^n \lambda^m \tag{8}$$

So then if we write C^m using equation (2) and the orthonormality of the eigenvectors we have

$$\mathbb{E}_z\langle z|C^m|z\rangle = \sum_i^N \lambda_i^m \mathbb{E}_z|\langle z|\phi_i\rangle|^2 = Tr(C^m) \tag{9}$$

where the last equality holds for any zero mean unit variance random vector from the expectation of quadratic forms.

$$Tr\mathbb{E}(z^tCz) = \mathbb{E}Tr(zz^tC) = Tr(C\mathbb{E}(zz^t) = Tr(C\Sigma) + \mu^tC\mu \tag{10}$$

Why should the dot product between a random vector and the eigenvector be 1? Surely that means on average an eigenvector completely overlaps with a random vector, which seems nuts? where have I messed up

References

.