

GP Write Up 6

October 3, 2018

Convolutions on the Kernel matrix

1. Convolution when missing observations

Lets say we have N points on a grid. Out of which M have been observed and W are missing. We fill the W missing points with very high variance normal distribution, i.e $\sigma_w \gg \sigma_m$. The kernel matrix can be written as

$$K_N = \begin{bmatrix} K_M & K_{MW} \\ K_{MW}^T & K_W \end{bmatrix} \quad (1)$$

where K_M is the co-variance matrix for all observed points and K_{MW} is the covariance matrix between observed and imaginary points.

and the noise matrix is:

$$D_N = \begin{bmatrix} \sigma_M^2 I_M & 0 \\ 0 & \sigma_W^2 I_W \end{bmatrix} \quad (2)$$

and the observations are $y = [y_M; y_W]^T$.

Hence we have:

$$(K_N + D_N)^{-1} y = v \quad (3)$$

we can use the block matrix inversion algorithm to see what this inverse leads to:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & -A^{-1}B(I - D^{-1}CA^{-1}B)^{-1}D^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & (I - D^{-1}CA^{-1}B)^{-1}D^{-1} \end{bmatrix} \quad (4)$$

where:

$$A = K_M + \sigma_M^2 I_M \quad (5)$$

$$B = K_{MW} \quad (6)$$

$$C = K_{MW}^T \quad (7)$$

$$D = K_W + \sigma_W^2 I_W = \sigma_W^2 (\sigma_W^{-2} K_W + I_W) \quad (8)$$

$$\lim_{\sigma_W \rightarrow \infty} D^{-1} = 0 \quad (9)$$

Therefore equation (3) becomes

$$(K_M + \sigma_M^2 I_M)^{-1} y_M = v_M \quad (10)$$

and

$$y_M = (K_M + \sigma_M^2 I_M) v_M \quad (11)$$

But the problem here is that in writing the kernel matrix as equation (1) we have lost the grid structure that leads us to a convolution, because the observations can be missing from any point in the grid. Not sure how to eliminate this problem.

From equation (11) the predictive mean for the entire grid will be:

$$\mu_L = K_{NM} v_M \quad (12)$$

2. Resolution to the above problem

Lets get to via an example. Consider an image that is 2×2 on the grid $(0, 0), (0, 1), (1, 0), (1, 1)$.

The co-variance matrix can be written as:

$$K_N = \begin{bmatrix} (0,0) & (0,-1) & (-1,0) & (-1,-1) \\ (0,1) & (0,0) & (-1,1) & (-1,0) \\ (1,0) & (1,-1) & (0,0) & (0,-1) \\ (1,1) & (1,0) & (0,1) & (0,0) \end{bmatrix} \quad (13)$$

from the previous write up

$$K' = \begin{bmatrix} k(-1,-1) & k(-1,0) & k(-1,1) \\ k(0,-1) & k(0,0) & k(0,1) \\ k(1,-1) & k(1,0) & k(1,1) \end{bmatrix} \quad (14)$$

and

$$v' = \begin{bmatrix} v_{1,1} & v_{1,0} \\ v_{0,1} & v_{0,0} \end{bmatrix} \quad (15)$$

$$y_N \approx K' * v' \quad (16)$$

now lets say we are missing the observation on the grid for $(1, 1)$, from equation (11) the co-variance matrix corresponding to actual observations is taken into account.

$$K_M = \begin{bmatrix} (0,0) & (0,-1) & (-1,0) \\ (0,1) & (0,0) & (-1,1) \\ (1,0) & (1,-1) & (0,0) \end{bmatrix} \quad (17)$$

To get to equation (11) via convolution:

define a new v''

$$v'' = \begin{bmatrix} 0 & v_{1,0} \\ v_{0,1} & v_{0,0} \end{bmatrix} \quad (18)$$

and convolve with K' but drop the last convolution i.e between the submatrix k'_4 and v'' .

$$k'_4 = \begin{bmatrix} k(0,0) & k(0,1) \\ k(1,0) & k(1,1) \end{bmatrix} \quad (19)$$

Similarly if $(0, 1)$ is missing from the grid then convolve v'' with K' and drop the second convolution.

$$v'' = \begin{bmatrix} v_{1,1} & v_{1,0} \\ 0 & v_{0,0} \end{bmatrix} \quad (20)$$

Similarly the process can be extended to multiple missing observations.

Can we do a **fancy zero padding** and get away with dropping convolutions?