Covariance Model A: This has 4 (symmetric) inputs to the neural network model of K.

$$[1 \quad \log(x'\Sigma y + \epsilon) \quad \log((x-y)'(x-y) + \epsilon)]$$

where

$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Covariance Model C: This has 11 (symmetric) inputs to the neural network model of K, and uses the (optimized) hyperparameter values. It is based on an example of a covariance function in section 4.2.2 (pg 89) of Rasmussen.

$$[1 \quad x'\Sigma y \quad (x'\Sigma y)^2 \quad \sigma_f^2 \exp[-\frac{1}{2l^2}(x-y)'(x-y)] \quad]$$

where

$$\Sigma = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

and $a = b = c = 1$. Now ideally, Σ above should be PD, and it would be nice if the neural network modeling was able to enforce $ac - b^2 > 0$, but alas, it can't (optimization with a quadratic constraint like this is in general hard).