**Section 4: Model fitting**

For finding the best model we consider seven kernels: Constant, Dot Product (DP), Matern, Pairwise, RBF, Rational Quadratic(RQ) and White kernel. We consider all pairwise addition and multiplication among all these kernels and calculate their BIC score. To represent the BIC score for different combinations of kernels, we use heatmap. The lower triangle of our BIC score heatmaps represent multiplication and upper triangle represent addition of two corresponding kernels. We keep the diagonal blank to partition the upper and lower triangle. If any combination of kernel does not work (probably not positive semi definite) we keep it blank as well. We set a cutoff of 500 for the BIC score. Any BIC score greater or equal to 500 is a high BIC score and we set the score 500. We will follow this interpretation of BIC score heatmaps for Branin, LDA and SVM parts.

* **Part 2: LDA and SVM data:**

We are considering the transformed data from the visualization part as BIC score for non-transformed data is not good. The BIC score heatmap for LDA data is shown below.



From the above heatmap we see the best model is Constant\*RQ (BIC score -115.41). We have found the best hyperparameter of this kernel is ConstantKernel(1.73\*\*2) + RationalQuadratic(alpha=0.0001, length\_scale=16.3). Note that the DP\*Pairwise BIC is blank indicating some error (probably not positive semi definite).

Next we show the BIC score heatmap for SVM data:



For SVM data, Matern\*RQ has the best performance (BIC score -47.06) and the the best hyperparameter of this kernel is Matern(length\_scale=1.47e+04, nu=1.5) \* RationalQuadratic(alpha=0.000402, length\_scale=0.572). Again DP\*Pairwise is blank.