Question 4

The ridge regression solves the problem:

and it has the optimal solution:

=

Let

)

So could be written as

Given ,

=

Define the Gram matrix to be the matrix with

will then equal to

Since produces vector with the element equal to , we just need the kernel function but not the feature vector to predict .