
Machine Learning Homework #2

Joe Ellis - jge2105

October 14, 2013

1 PROBLEM 1

2 PROBLEM 2

A stationary kernel is some kernel such that the kernel result is translation invariant, we denote this as $K(x, z) = K_s(x - z)$. We have two examples of stationary kernels for this problem. The first example is the exponential kernel, $K_e(x, z) = e^{(-\frac{\|x-z\|}{\theta})}$, and the second example is the gaussian kernel, which is $K_g(x, z) = e^{(-\frac{\|x-z\|^2}{\theta})}$.

We will show that they are mercer kernels based on the definition given in the lecture slides, that a mercer kernel is a kernel which is kernel that takes two inputs and outputs a scalar via denoted as,

$$k(x_1, x_2) = \langle \phi(x_1), \phi(x_2) \rangle = \begin{cases} \phi(x_1)^t \phi(x_2) & \text{for finite } \phi \\ \int_t \phi(x_1, t) \phi(x_2, t) & \text{otherwise} \end{cases} \quad (1)$$

Both of these are mercer kernels, because they both output scalar values, for both conditions. Let's assume that ϕ is finite, and then we know that on the top of the exponential function we have a magnitude which is scalar divided by a scalar parameter θ which yields some scalar output value. We also, know that the integral of the kernel provided is bounded due to the nature of the gaussian and exponential kernel, and therefore $\int_t \phi(x_1, t) \phi(x_2, t)$, is also bounded. Therefore, both are mercer kernels.

Now let's show that both of these are stationary kernels. We will translate both of the elements by adding some arbitrary variable k , which is of the same dimensionality of the

inputs x_1 and x_2 . Therefore, for any k we want to show that the kernels are stationary under sum and product. First, assume we sum the kernels together,

$$K_{sum}(x, z) = K_e(x, z) + K_g(x, z) \quad (2)$$

$$= e^{(-\frac{\|x-z\|}{\theta})} + e^{(-\frac{\|x-z\|^2}{\theta})}, \quad (3)$$

$$= K_{se}(x - z) + K_{sg}(x - z). \quad (4)$$

$$(5)$$

Since both values that we receive in the summation depend only on the lag vectors, then we find that they are stationary under summation.

Now let's show that these are also stationary kernels under multiplication. For multiplication we have,

$$K_{mult}(x, z) = K_e(x, z)K_g(x, z) \quad (6)$$

$$= e^{(-\frac{\|x-z\|}{\theta})} e^{(-\frac{\|x-z\|^2}{\theta})}, \quad (7)$$

$$= e^{(-\frac{\|x-z\| + \|x-z\|^2}{\theta})} \quad (8)$$

$$(9)$$

Since the exponent of the above multiplication of the kernels also only depends on the value $x - z$, we have shown that these are also stationary under multiplication.

For the second part of this problem, we will explicitly find the ϕ function such that $K(x_1, x_2) = \phi(x_1)^t \phi(x_2) = e^{(-\|x_1 - x_2\|^2)}$.

$$K(x_1, x_2) = \phi(x_1)^t \phi(x_2) \quad (10)$$

$$= e^{-\|x_1 - x_2\|^2} \quad (11)$$

$$= e^{-(x_1 - x_2)(x_1 - x_2)} \quad (12)$$

$$= e^{-(x_1^2 + x_2^2 + 2x_1x_2)} \quad (13)$$

3 PROBLEM 3

For this problem we calculated the gram matrix for a given kernel $f(x, \lambda, k)$. We define $f(x, \lambda, k)$ as 0 for $x < 0$, and $f(x, k, \lambda) = (\frac{k}{\lambda})(\frac{x}{\lambda})^{k-1}(e^{-(\frac{x}{\lambda})^k})$. We define our kernel function $K(x_i, x_j) = F(x_i^t x_j, \lambda = 0.75, k = 3)$, where $F(x, \lambda, k) = \int_0^x f(x, \lambda, k) dx$.

When we integrate $(f(x, \lambda, k))$, we get for the indefinite integral,

$$\int f(x, \lambda, k) dx = F(x, \lambda, k) = -e^{(\frac{x}{\lambda})^k}. \quad (14)$$

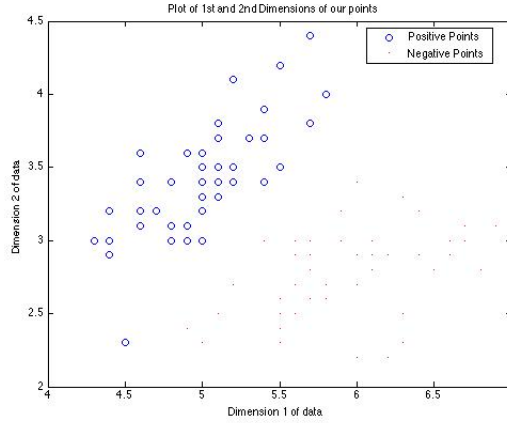


Figure 1: Plot of the 1st and 2nd dimensions of the dataset

Therefore, using this knowledge and computing the definite integral between two points, we created a Matlab script to calculate the desired gram matrix. The gram matrix can be seen here,

$$\begin{pmatrix} 0.0698 & 0.0456 & 0.0104 & 0.0072 & 0.0635 \\ 0.0456 & 0.0420 & 0.0043 & 0.0061 & 0.0191 \\ 0.0104 & 0.0043 & 0.0053 & 0.0007 & 0.0080 \\ 0.0072 & 0.0061 & 0.0007 & 0.0009 & 0.0039 \\ 0.0635 & 0.0191 & 0.0080 & 0.0039 & 0.2478 \end{pmatrix}$$

This is the gram matrix for the points given in problem 3.

4 PROBLEM 4

Here we create an object recognition system using SVM for classification. Multiple different SVM kernels were chosen and evaluated, and all of the kernels chosen were able to perform well with the proper parameters. We performed cross-validation across the data-set using a 50-50 train and test split for the sample data points. We performed cross-validation with 10 random splits for each parameter to show the performance of the SVM with each different kernel. Given the proper parameters we were able to get 100% accuracy using each of the described kernels, and this is because the data provided to us was seperable. Figure 4 shows the positive and negative samples from our dataset plotted with respect to their first two dimensions. We can see that they are seperable through a linear svm.

For classification we used linear, RBF, and polynomial SVM with different values for the cost variable (C), and the number of degrees for the polynomial function and the size of sigma in the RBF. Each of these paramters effected the classification accruacy, but proper parameters for each kernel scheme could be found. The results of a different

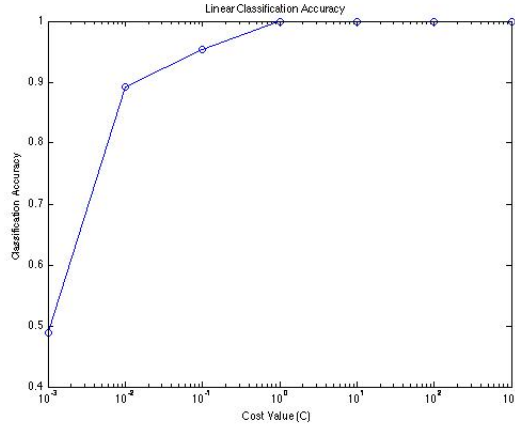


Figure 2: Linear SVM classification accuracy for different parameters

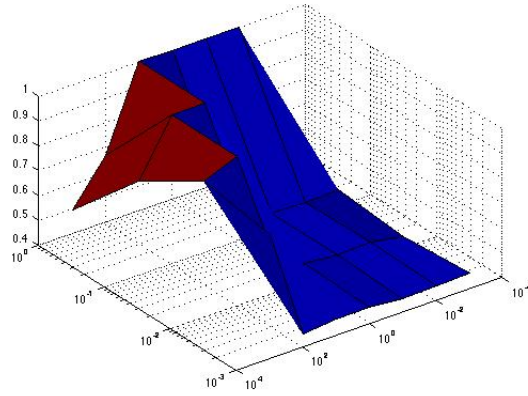


Figure 3: RBF SVM classification accuracy for different parameters

parameters can be seen in the upcoming figures.

As you can see, we have found parameters for each type of kernel that achieves maximum accuracy. However, the kernel that in general used the least amount of support vectors, and was quickest in classification was the linear SVM. Therefore, for this particular classification task I would recommend using the linear SVM with $C = 10$. However, if the dataset is more difficult and not linearly separable we would instead need to focus on using a more flexible kernel function such as the RBF.

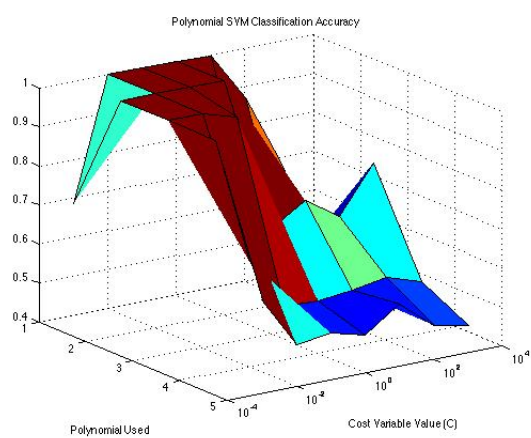


Figure 4: Polynomial SVM classification accuracy for different parameters