Assignment 1

Dhaval Parmar

June 7, 2020

1 Valid Kerenel

1.1

Question 1:

a)

- If $A_{d \times d}$ is a positive semidefinite matrix, then the function $k: R^d \times R^d \to R$ given by $K(x, y) = x^T A y$ is a valid kernel. Here $x, y \in R^d$ are d-dimensional vectors.

Proof:

- ~~let~~ Since A is PSD we can write it in the form $A = BB^T$.
- For every $N \in \mathbb{N}$, and every choice $x_1, \ldots, x_N \in \mathbb{R}^d$ we form the matrix $k = (k_{ij})$ where

$$k_{ij} = k(x_i, x_j) = x_i^T A x_j$$

so for every $c \in \mathbb{R}^N$

$$\sum_{i=1}^{N} \sum_{j=1}^{N} c_i c_j k_{ij} = \sum_{i=1}^{N} \sum_{j=1}^{N} c_i c_j x_i^T A x_j$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} c_i c_j (B^T x_i)^T (B^T x_j)$$

$$= \| \sum_{i=1}^{N} c_i (B^T x_i) \|^2$$

$$\geq 0$$

Here $\phi(x) = B^T x$.

- So k is valid kernel.

Claim: A is PSD if and only if $A = BB^T$ for some real matrix B.

Proof: let $A = BB^T$ for any vector v.

$$v^T A v = v^T B B^T v = \| B^T v \|^2 \geq 0$$

so A is PSD.

- Now let A is PSD. let $AV = V\Lambda$ be the eigen decomposition of A and ~~B =~~ $B = V\sqrt{\Lambda}$ where $\sqrt{\Lambda}$ is the diagonal matrix with entries $(\sqrt{\Lambda})_{ii} = \sqrt{\lambda_i}$. The matrix exists since the eigen values are non negative.

$$BB^T = V\sqrt{\Lambda}\sqrt{\Lambda}V^T = V\Lambda V^T = A.$$

1.2

1 No

2 Yes

3 No

4 No

5 No

2 Support Vector Machines

2.1

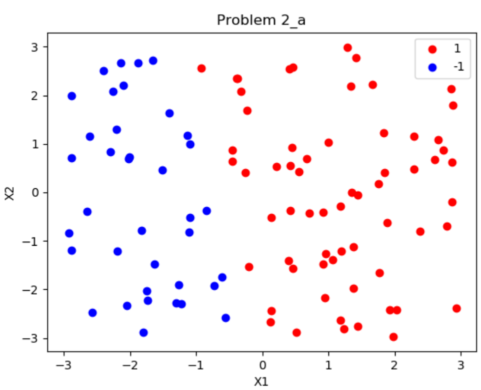

Figure 1: Train Data

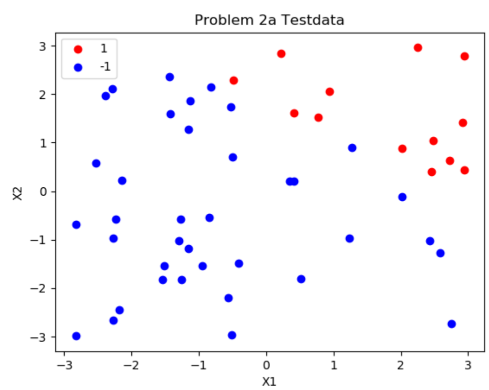

Figure 2: Test Data

2.2

W = [1.12, 0.207]
b = 0.6241
c = 0.04
Train Accuracy = 100%
Test Accuracy = 76%

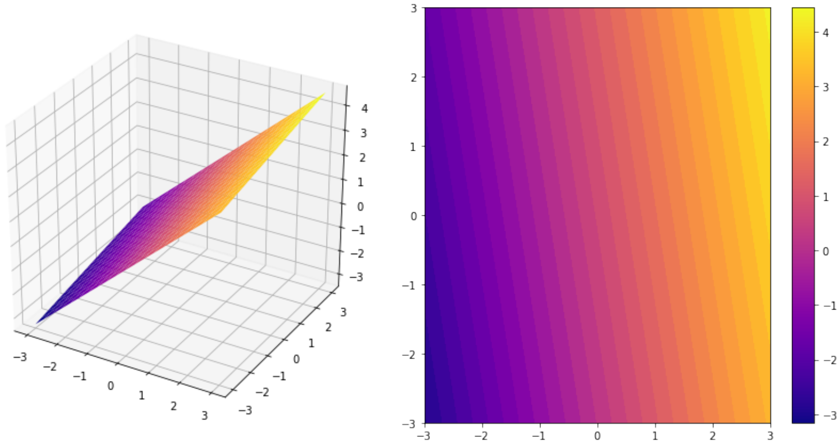

Figure 3: Learned Function Value

2.3

2.3.1

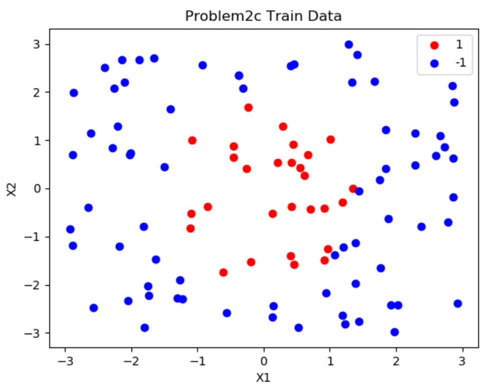

Figure 4: Train Data

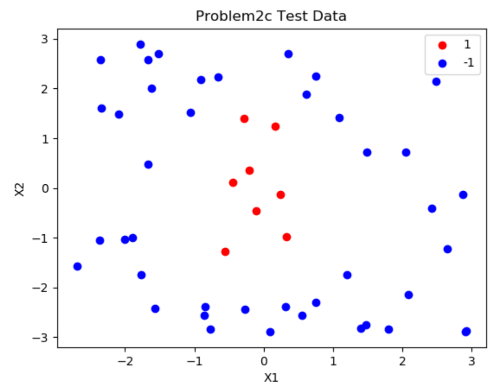

Figure 5: Test Data

2.3.2

W = [5.03e-11, 1.43e-12]
b = -4.2148e-10
c = 3
Train Accuracy = 72%
Test Accuracy = 84%

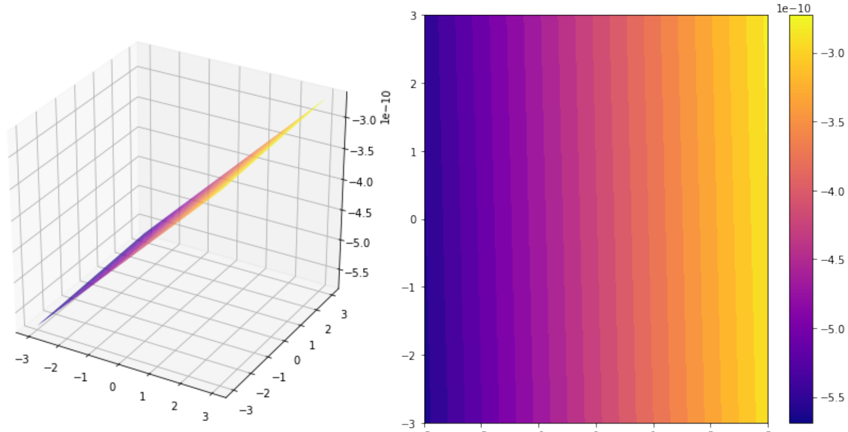

Figure 6: Learned Function Value

2.4

2.4.1

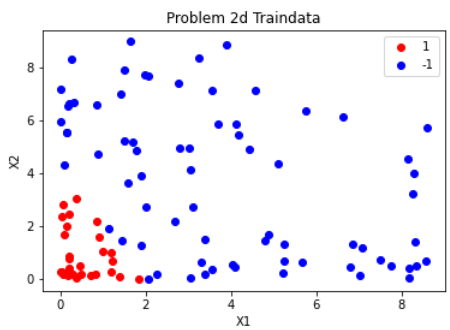

Figure 7: Train Data

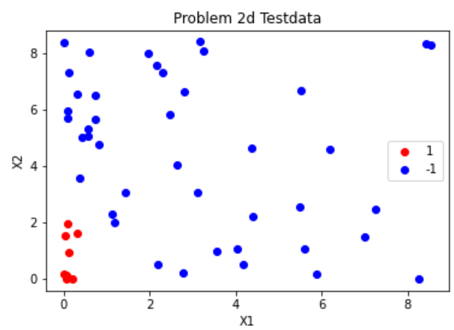

Figure 8: Test Data

2.4.2

W = [-1.96e-4, -1.68e-4]
B = 5.8765e-4
C = 1.2e-6
Train Accuracy = 96%
Test Accuracy = 94%

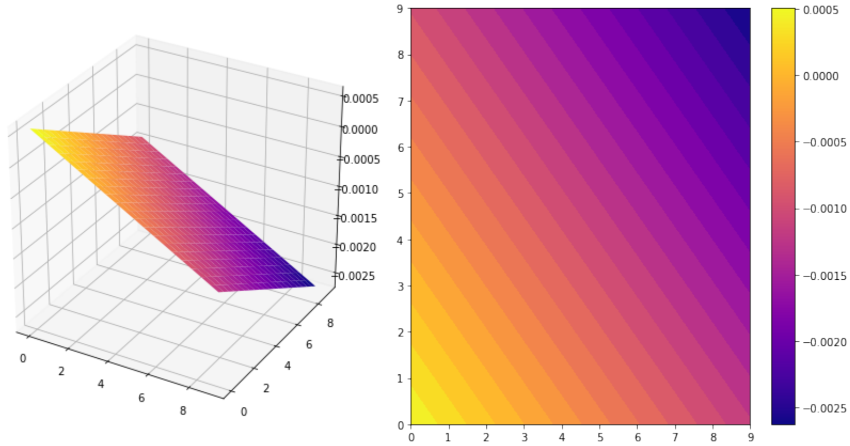

Figure 9: Learned Function Value

2.4.3

Here we can see that performance is increased compare to problem 2.3. From the plot of data we can see that data is linearly separable in this case while it is not linearly separable in problem 2.3. So hyper plan can better fit in case of problem 2.4. So here we can see that data that is not linearly separable can be made linearly separable and that makes classification easy.

2.5

2.5.1

Kernel Function $K(x, y) = (a + bx^T y)^d$
a = 1
b = 1
d = 2
c = 5 (Upper limit on alphas)
Train accuracy : 91%
Test accuracy : 98%

3

3.1

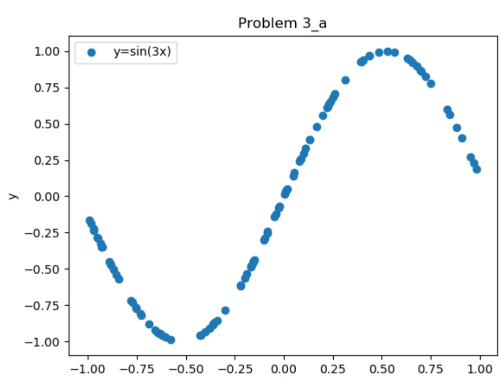

Figure 10: Train Data

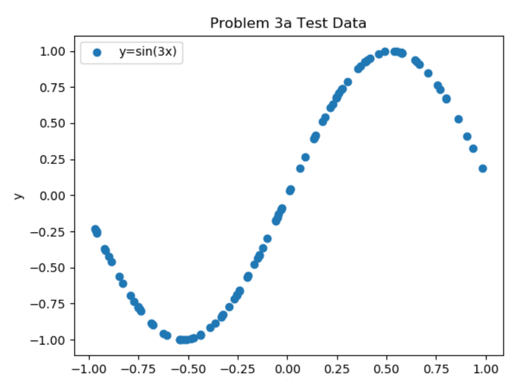

Figure 11: Test Data

3.2

Mean Square Error : 0.1563935

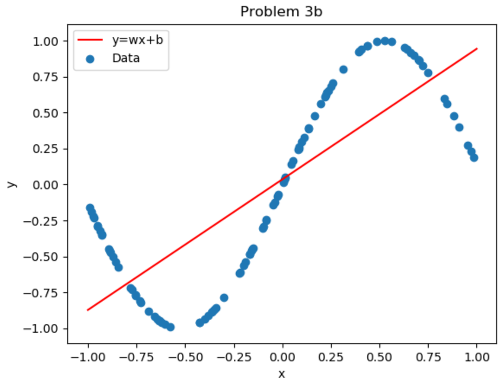

Figure 12: Learned Function

3.3

MSE for k=1 is 0.15639355625140347
MSE for k=2 is 0.15116877830522277
MSE for k=3 is 0.0030563262941716274
MSE for k=4 is 0.0029808503600035003
MSE for k=5 is 9.640933389612657e-06
MSE for k=6 is 9.3538611350001e-06
MSE for k=7 is 1.0230999599102144e-08
MSE for k=8 is 9.793018422099174e-09
MSE for k=9 is 4.728021137960336e-12
MSE for k=10 is 4.093252683490134e-12

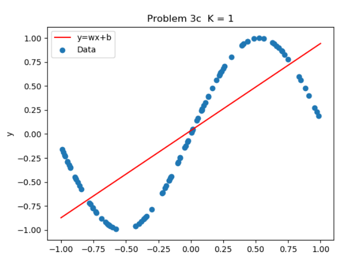

Figure 13: k = 1

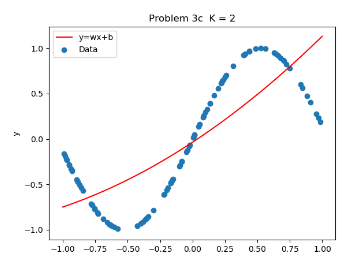

Figure 14: k = 2

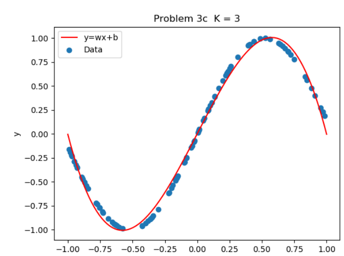

Figure 15: k = 3

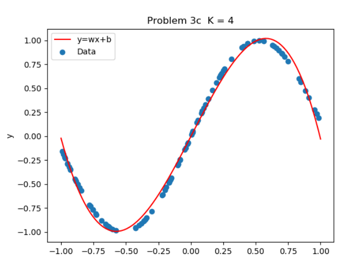

Figure 16: k = 4

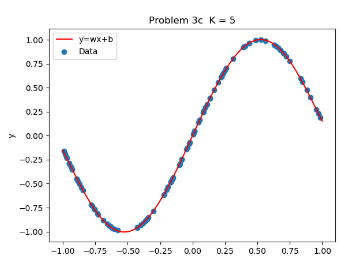

Figure 17: k = 5

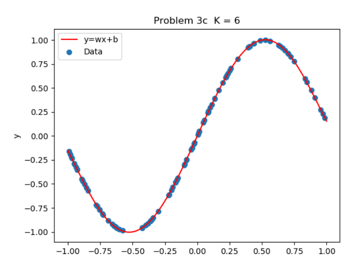

Figure 18: k = 6

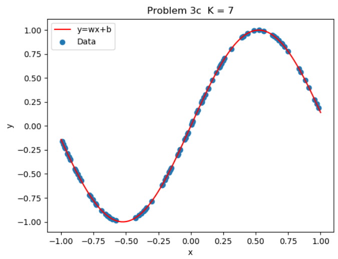

Figure 19: k = 7

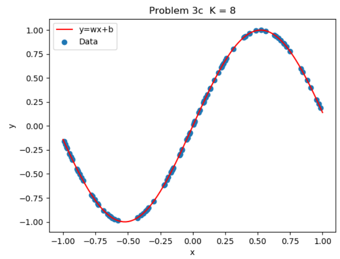

Figure 20: k = 8

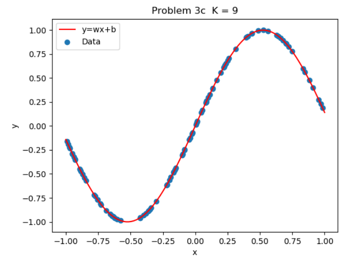

Figure 21: k = 9

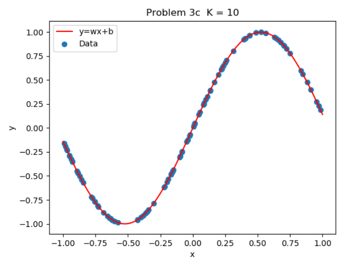

Figure 22: k = 10

3.4

Kernel function used is Rational Quadratic Kernel.
$k(x, y) = 1 - \frac{||x-y||^2}{||x-y||^2+c}$

c = 0.065
MSE for train set = 8.267447579118345e-12
MSE for test set = 0.4494755728479644

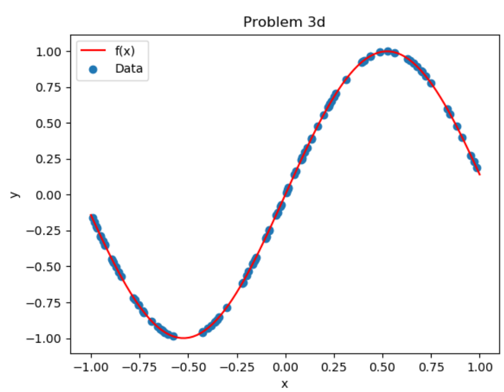

Figure 23: Learned Function

4

4.1

Sum of all entries of D = 0

4.2

$\sum_{i=1}^{d} d_{\text{i}} = 0$

4.3

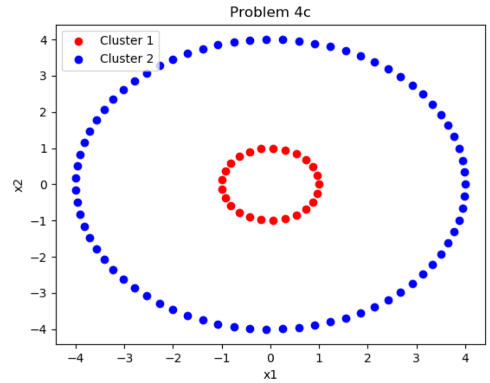

Figure 24: Clusters

4.3.1

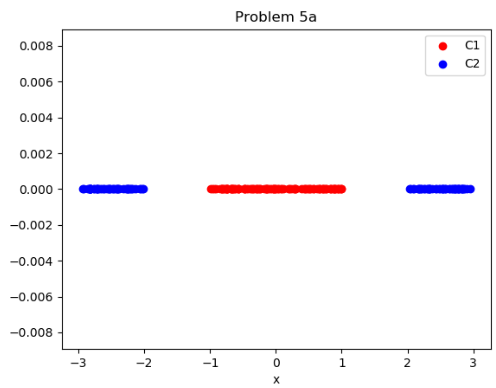

Figure 25: Data

4.3.2

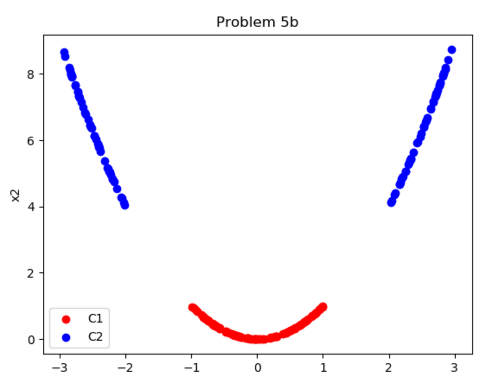

Figure 26: Data

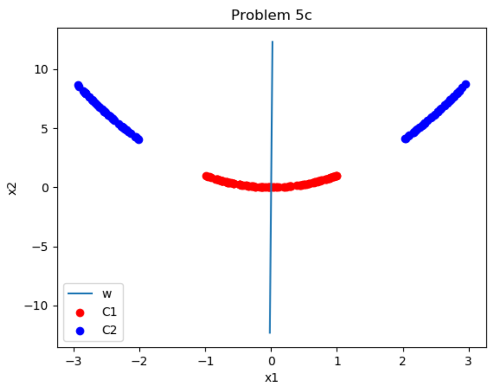

Figure 27: Direction W

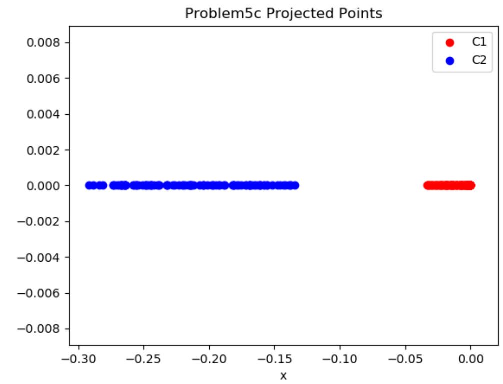

Figure 28: Projected Data on 1D

4.3.3

4.3.4

Classification Accuracy = 100%

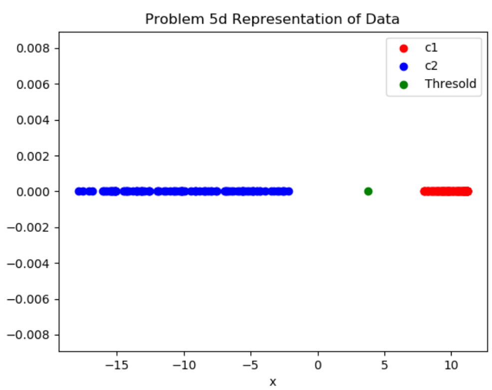

Figure 29: Projected Data on 1D

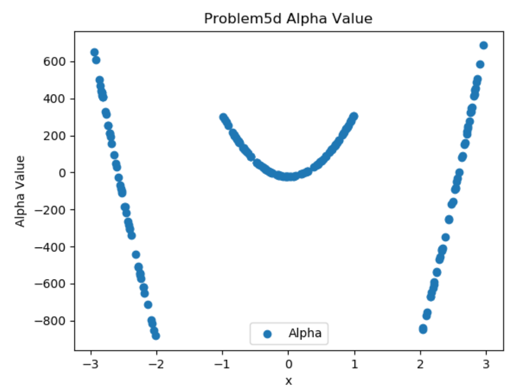

Figure 30: Alpha Values