Advanced Kernel Methods - Assignment 2

Ragja Palakkadavath - SC18M003

January 2019

1. **Solution:**
   Sobolev Space is an example of an RKHS. $f : [0, 1] \to \mathbb{R}$ with f(0)=f(1)=0.
   The norm in this space is defined by : $\| f_H \|^2 = \int (f^{'}(x))^2 dx = \int \omega^2 \mid F(\omega) \mid d\omega$

2. **Solution:**
   A Hilbert space H becomes a RKHS F if and only if every point evaluation functionals L_{x_i} that is defined on H are bounded and linear. For L_{x_i} in H to be bounded,
   $\forall\ x \in X$, $\|L_x(f)\| = \|f(x)\| = |f(x)| \leqslant \|L\|\,\|f\|$, which means,
   $|f(x)| \leqslant c\,\|f\|\ \ \forall\ c > 0$
   Checking whether any point evaluation functional in $L_2[0, 1]$ is bounded and linear or not:
   Let $f = x^n \in L_2[0, 1]$, then $\|f\|^2 = \langle f, f\rangle = \int_0^1 (f(x))^2 dx$,
   $\|f\| = \left[\int_0^1 (f(x))^2 dx\right]^{1/2} = \left[\int_0^1 x^{2n} dx\right]^{1/2} = \frac{1}{\sqrt{2n+1}}$
   $\therefore \lim_{n\to\infty} \frac{1}{\sqrt{2n+1}} = 0$
   At $x = 1, f(1) = (1)^n = 1$ and $\|f\| \to 0\ \forall n$
   $\therefore$ It is not possible to find a c that can satisfy the equation: $|f(x)| \leqslant c\,\|f\|\ \ \forall n$, which means that f is not bounded. Therefore, $\|L_x(f)\|$ is also not bounded. Hence, $L_2[0, 1]$ is not a RKHS.

3. **Solution:**

   (a) CVXOPT package was used for solving the dual objective function.

   (b) Plot of Decision Boundary: Positive points are in blue, negative points are in red and the support vectors are given in green.

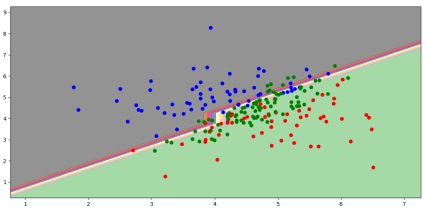

Figure 1: Data1 Plot

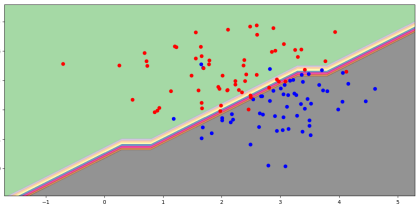

Figure 2: Data2 Plot

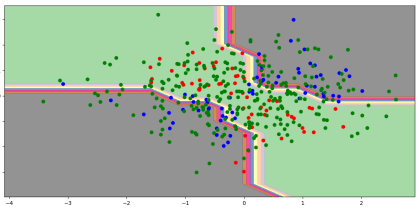

Figure 3: Data3 Plot

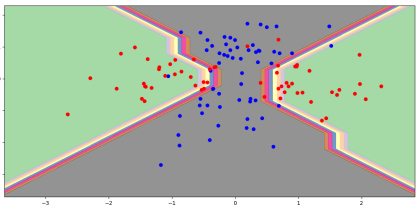

Figure 4: Data4 Plot

(c) (i) Data 1
Duality GAP is: 7.02705383925488e-06
Dual Objective Function value is: -2633.3636115724257
Primal Objective Function value is: -2633.3636152995996
Dual Slack variable value is 4.760232566733569e-11
Primal Slack variable value is 4.5149164478011057e-10
Value of bias is 4.187959188283016
Accuracy obtained is 1.0
C obtained after cross validation is 10000

(ii) Data 2
Duality GAP is: 630895479.0425467
Dual Objective Function value is: -530652373436.9928
Primal Objective Function value is: -531283268915.8739
Dual Slack variable value is 50903.99907749032
Primal Slack variable value is 4.6760843213627395e-06

Value of bias is 839001.3711372843
Accuracy obtained is 0.8916666666666667
C obtained after cross validation is 10000

(iii) Data 3
Duality GAP is: 0.001303492497939946
Dual Objective Function value is: -8691.38706189985
Primal Objective Function value is: -8691.388365392348
Dual Slack variable value is 1.9979245279348966e-09
Primal Slack variable value is 7.744077927335914e-10
Value of bias is -3.5350543436941995
Accuracy obtained is 0.975
C obtained after cross validation is 10000

(iv) Data 4
Duality GAP is: 1260879463.6238523
Dual Objective Function value is: -1140988825551.778
Primal Objective Function value is: -1142249705015.0503
Dual Slack variable value is 2041.8622520231381
Primal Slack variable value is 6.455173060607306e-09
Value of bias is 40.871262642799344
Accuracy obtained is 0.9416666666666667
C obtained after cross validation is 1

(v) f/w values and α values for all the data above are present in the attached csv files

4. **Solution:**

(a) Non linear plot for data set H

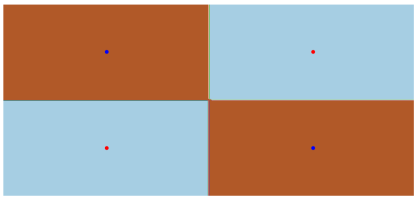

Figure 5: Data set H Plot

(b) SVM Classification on Data3new, Data4new parameters of the model

(i) Data3new
Duality GAP is: 0.0005041441728120745
Dual Objective Function value is: -14030.040729550727
Primal Objective Function value is: -14030.041199679468
Dual Slack variable value is 3.07883921575528e-09
Primal Slack variable value is 4.4797550476859113e-10
Value of bias is 15.23387381419978
Accuracy obtained is 0.9
C obtained after cross validation is 10000

(ii) Data4new
Duality GAP is: 228830626.18760592
Dual Objective Function value is: -843271662168.2017
Primal Objective Function value is: -843500492794.3335
Dual Slack variable value is 721.0231281399314
Primal Slack variable value is 6.0718521594421114e-05
Value of bias is 0.4907212272071761
Accuracy obtained is 0.8916666666666667
C obtained after cross validation is 0.01

(iii) f/w values and α values for all the data above are present in the attached csv files

(c) SVM Classification Plots

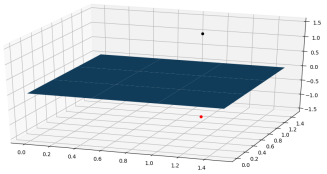

Figure 6: Data set H new Plot

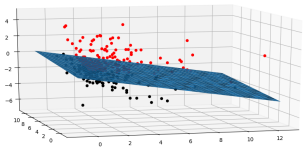

Figure 7: Data set Data 3 new Plot

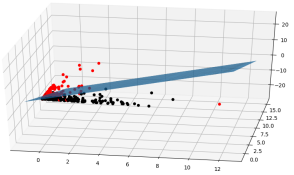

Figure 8: Data set Data 4 new Plot

(d) Following are the values obtained after calculating $\tilde{f}, \tilde{f}_3, \tilde{f}_4$ and $f^{\tilde{new}}, f_3^{\tilde{new}}, f_4^{\tilde{new}}$ for H, Data3 and Data4:

$$| \tilde{f}(x_i) - f^{\tilde{new}}(x_i) | = 0$$

$$| \tilde{f}_3(x_i) - f_3^{\tilde{new}}(x_i) | = 8.03965312234561e-12$$

$$| \tilde{f}_4(x_i) - f_4^{new}(x_i) | = 4.70593773486757e-13$$

Therefore it can be concluded that they are equal in values.

5. **Solution:**
Given $\tilde{f}(x) = \sum_{i=1}^{N} \alpha_i \langle x_i, x \rangle + b$, where $\tilde{f}(x)$ belongs to RKHS F.
Also given, $h(x) = \alpha_1 x_1 + \alpha_2 x_2 + .. + \alpha_n x_n + b$, which is an equation in $\mathbb{R}^n$.

$\tilde{f}(x)$ represents a hyperplane which is formed using a linear reproducing kernel(input function itself is of the linear form).

Since the RKHS F and the euclidean space $\mathbb{R}^n$ can be related by a one to one mapping, the expressions $\tilde{f}$ and h are same.

6. **Solution:**

(a) CVXOPT package was used for solving the dual objective function.

(b) Plot of Decision Boundary: Positive points are in blue, negative points are in red and the support vectors are given in green.

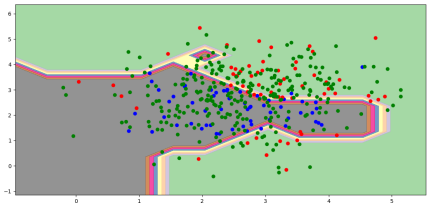

Figure 9: Data5 Plot

(c) Data 5 Parameters
Duality GAP is: 0.0013369243892240347
Dual Objective Function value is: -17712.314243437075
Primal Objective Function value is: -17712.315580361465
Dual Slack variable value is 2.0682470422925043e-06
Primal Slack variable value is 3.1454590811904395e-09
Value of bias is -0.927034535470621
Accuracy obtained is 0.9
C obtained after cross validation is 10000
Best kernel is Hyperbolic

(d) f/w values and α values for all the data above are present in the attached csv files

(e) Cross validation was performed among linear, polynomial, gaussian and hyperbolic kernels and the kernel that gave the best accuracy measure was chosen as the preferred one.

(f) KKT Complementary conditions have been satisfied by all the data points in Data 5.