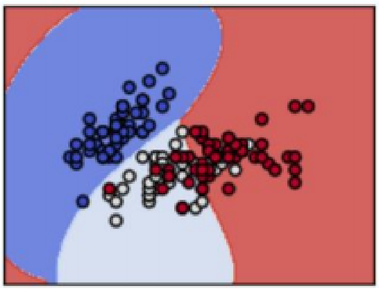
CSE/ECE 343/543: Machine Learning
Assignment-2 SVM

Max Marks: 100 (Programming:70, Theory:30) **Due Date**: 10/10/2019, 11:59PM

Instructions

- Keep collaborations at high level discussions. Copying/Plagiarism will be dealt with strictly.
- Late submission penalty: As per course policy.
- Your submission should be a single zip file **name_HW2.zip**. Including only the **relevant files** arranged with proper names. A **.pdf report** explaining your codes with relevant graphs and visualization and theory questions.
- Remember to **turn in** after uploading on google classroom.
- **Bonus** of 10 points for submissions before **6 hours of deadline.**
- Resolve all your doubts from TA's in their office hours **two days before the deadline.**

1. (5+10+10+20 = 45 points) **Exploring dataset and kernels**
   - You are given 5 different 2 dimensional datasets. Some datasets are noisy, unbalanced etc. Explore the datasets, plot them and write your observations and findings of the datasets.
   - For each dataset, write your own kernel (if required) to make them linearly separable. Plot the datasets with decision boundaries corresponding to those kernels, something like this:



     Explain the choice of kernels.
   - Use outlier removal techniques to remove outliers in the datasets. Plot the outlier-removed datasets also.
   - For dataset 4 and 5. Split the datasets into 80-20 split and then, use the training datapoints to fit a linear SVM and predict the output classes on both the train and test set. Do the same for an RBF kernel.
     Note: You are only allowed to use sklearn.fit function but have to implement the predict function on your own. Also for RBF kernel, you have

to write the kernel function on your own from scratch. (link). Compare the accuracy of your model with the Sklearn SVM prediction function.

2. (25 points) **SVM on CIFAR-10 dataset**

   Download the CIFAR-10 dataset (use the python version of the dataset for this question and the folds given in the link). Apply the following methods to classify the dataset, for both one-vs-all and one-vs-one classification methods.

   1. SVM with no kernel
   2. SVM with RBF kernel
   3. SVM with quadratic polynomial kernel

   Note: You should provide the weights for all 30 trained SVM models (5 folds * 3 kernels * 2 (1-v-1/1-v-all)).

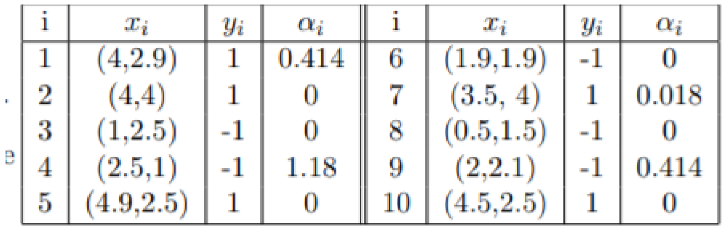
   Which of the above kernels performs the best. Cite your reasons with apt visualizations, plots and tables (include confusion matrices, accuracy reports and ROC curves).
   You are allowed to use the inbuilt sklearn functions for the SVM implementation and the inbuilt kernels in sklearn (link).

---

3. (10 points) Can you model the XOR operator using an SVM? Justify

4. (10 points) Consider a Kernel K such that K(x, x') = $(1 + x^T x')^2$. Also, consider x to be 2-dimensional with features x_1 and x_2 aka x = $[x_1, x_2]^T$. What is the transformation $\phi(x)$ in this case? How many features does $\phi(x)$ contain? [Hint: K(x, x') is the dot product of $\phi(x)$ and $\phi(x')$ by definition]

5. (10 points)



i	x_i	y_i	α_i	i	x_i	y_i	α_i
1	(4,2.9)	1	0.414	6	(1.9,1.9)	-1	0
2	(4,4)	1	0	7	(3.5, 4)	1	0.018
3	(1,2.5)	-1	0	8	(0.5,1.5)	-1	0
4	(2.5,1)	-1	1.18	9	(2,2.1)	-1	0.414
5	(4.9,2.5)	1	0	10	(4.5,2.5)	1	0

   Consider the training data samples and corresponding Lagranges multipliers learned from them, in the table,

   - What is the b for the SVM?
   - Identify the support vectors.
   - Compute w and classify point (3,3)

6. (BONUS – 10 points)
   Consider the RBF Kernel which maps data to an infinite dimensional space given by K(u, v) = $exp(-\gamma \parallel u - v \parallel^2)$
   Further, assume u and v to have only 1 feature i.e. u and v are real numbers.
   What is $\phi(u)$ in this case? What can be said about the coefficients of higher order terms of u (such as $u^{1000000}$) present in $\phi(u)$? [Hint: Since RBF Kernel maps data to an infinite dimensional space, $\phi(u)$ is an infinite dimensional feature vector. Use Taylor Series expansion for $exp(2uv)$ and try to separate out terms for u and terms for v to obtain the transformation ϕ. You might want to use γ as 1 for simplicity]