Course: Math 6397 Spring 2015

Instructor: Robert Azencott

Lecture Date: 2/19/2015

Notes prepared by Renat Tatarin

Draft 1

Project 2:

Using your 2 data sets from project 1 for which you performed LDA, now perform Support Vector Machine (SVM) classification. If your results for a data set exceeded 90% using LDA, consult with Dr. Azencot about possibly choosing another data set for analysis as we are interested in improving our results with SVM.

SVM Classification:

Suggested Software Package → SVM Light, e1071, and many others that can be found online

Website Suggestions for SVM Light:
R: http://www.inside-r.org/packages/cran/klaR/docs/svmlight
Matlab: http://mex-svm.sourceforge.net/
Mathematica: http://malchiodi.di.unimi.it/software/svMathematica/

- Use 2 different kernels:
  - Polynomial kernels of degree $l \in \{2,3,4\}$: $K(x,y) = \left[1+ < x,y >_{\mathbb{R}} >\right]^l$
  - Gaussian kernel: $K(x,y) = \exp\left(\frac{-|x-y|^2}{\sigma^2}\right)$
- For the Gaussian kernel, you will choose σ
  - Report all classification performance for various σ you chose. You can explore the range of σ values by successive dichotomies
  - For $\sigma \ll mean|x-y|$, good training performance, but poor testing performance will be achieved
  - For $\sigma \gg mean|x-y|$, the program will not learn well

- Construct a "Confusion (or Error) Matrix" for the training set and another confusion matrix for the test set
- Compare the two to evaluate the generalization capacity of the SVM classifier

		SVM Classification	
		$+1$	-1
True Class	$+1$	True Positive	False Negative
	-1	False Positive	True Negative

Separator: $Sep(x)$

We would like to find an explicit separator in the Hilbert space H such that the data, after being transformed into a higher dimension, can be separated into classes -1 and +1 by a hyperplane. For N cases with k attributes we have the attribute vectors $X(i) \in \mathbb{R}^k, i \in N$. Then

$$w = \sum_{i=1}^{N} \alpha_i Z_i \text{ where } \mathbb{R}^k \ni X(i) \to Z_i \in H \text{ (Hilbert)}$$

Here, α_i are the support vector coefficients and therefore w is a linear combination of these support vectors. They will either be given after running the program or you need to call for them. Many will be equal to 0 and may not be shown.

Case i is a support vector if and only if $|\alpha_i| > 0$. For some of the cases, you will have $\alpha_i < 0$ which represents the -1 class and for others you will have $\alpha_i > 0$ which represents the $+1$ class

If $j \in S =$ support vectors remaining after all $\alpha_i = 0$ have been eliminated, then $1 \leq S \leq N$. For robust classification: $S \ll N$.

Then the explicit $Sep(x)$ will be:

$$Sep(x) = \sum_{j=1}^{N} \alpha_j K\left[X(j), x\right] + b$$

The larger the α_i, the harder it is to classify that corresponding case. You should look at at least 10 support vectors having coefficients α_i with the highest absolute values and examine their corresponding cases to see if there are any errors in the data. This should look something like this:

$$|\alpha(1)| \geq |\alpha(2)| \geq |\alpha(3)| \geq \ldots \geq |\alpha(j)|$$

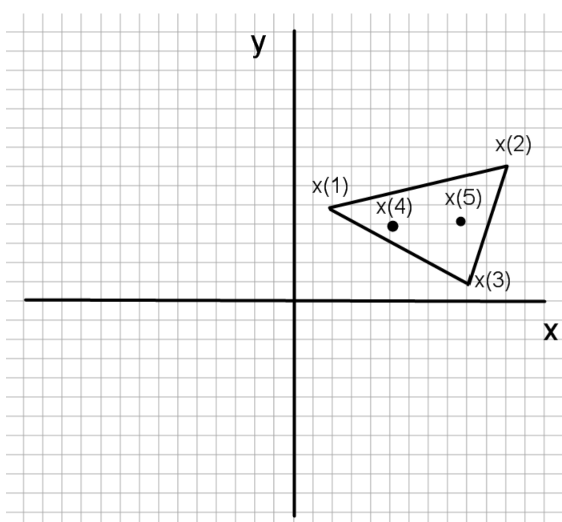
Be careful to keep track of the corresponding cases when ranking the support vectors.

Toy Problem: To be included in project 2

Construct a data set with the following information:

- $N = 5$ cases $X(i) \in \mathbb{R}^k, i \in N$
- $k = 2$ attributes
- 2 classes
  - $x(1), x(2), x(3)$ in class 1
  - $x(4), x(5)$ in class 2
- $2 \leq ||x(i)|| \leq 4, i \in$ class 1
- $||x(i) - x(j)|| \geq 1, i \neq j$
- $||x(4)|| < 1.5$
- $||x(5)|| \leq 1.5$
- $||x(4) - x(5)|| \geq 1$

It should look somewhat like this; one class forming a triangle around the other:



Perform separation with SVM

- Polynomial kernel of degree $k \in \{2, 4\}$
- Gaussian kernel
- LDA

You will see that the best separator is a polynomial kernel of degree 2.