Machine Learning from Data – IDC
HW5 – Support Vectors Machine

***** This assignment can't be submitted in pairs - individuals only.**

In this assignment you will use Weka's implementation of an SVM classifier (called SMO). First you will divide the data to training and test set. Then you will select the best kernel according to the TPR and FPR (with $\alpha = 1.5$). In addition, you will plot the result of each kernel in a ROC graph in excel. After selecting the best kernel, you will try different values for the parameter C (the slack regularization) and once again you will plot the result for each value in a ROC graph. Finally, you will answer some theoretical questions regarding kernel functions and Lagrange multipliers (at the end of the document).
The grades for this assignment constructs from 46% on the programing part and 54% for the theoretical part (6 Pts per question).

Finding the best kernel:

1. Divide the data to training and test set – 80% training and 20% test.
2. For each kernel, build the SVM classifier on the training set using the SMO WEKA class.
3. Calculate & print to the console the TPR and FPR on the test set.
4. Select the best kernel according to the best αTPR-FPR (with $\alpha = 1.5$).
5. The possible types for the kernel are:
    a. Polynomial Kernel – with the following degrees $\{2, 3, 4\}$
    b. RBF Kernel – with the following gamma values $\left\{\frac{1}{200}, \frac{1}{20}, \frac{1}{2}\right\}$
6. In the attached excel file, fill the TPR and FPR for each kernel in the appropriate yellow cells. Add a parallel line (to the line from the origin, which is already there) that shows the best kernel on the graph.

Finding the best C value (the slack regularization):

1. For the selected kernel, try different C values.
2. For each C value, build the SVM classifier with the selected kernel on the training set using the SMO WEKA class.
3. Calculate & print to the console the TPR and FPR on the test set.
4. Examine C values that are all the combinations $\left\{10^i * \frac{j}{3}\right\}$, where $i = \{1, 0, -1, -2, -3, -4\}$ and $j = \{3, 2, 1\}$

5. In the attached excel file, fill the selected kernel in the gray cell and fill the TPR and FPR for each C value in the appropriate green cells. The ROC graph will update automatically. Does your graph make sense?

In order to do the above you need to first install WEKA:

1. See instructions in HW1.

Prepare your Eclipse project:

1. Create a project in Eclipse called HomeWork5.
2. Create a package called HomeWork5.
3. Move the SVM.java and MainHW5.java that you downloaded from the Moodle into this package.
4. Add WEKA to the project (see instructions in HW1).

The following methods are mandatory methods and you can't change their signatures, but you can add additional methods:

1. void setKernel: Setting the Weka SMO classifier kernel.
   a. Input: Kernel object.
2. void setC: Setting the C value for the Weka SMO classifier.
   a. Input: double.
3. double getC: Getting the C value for the Weka SMO classifier.
   a. output: double.
4. Int[] calcConfusion: Calculate the TP, FP, TN, FN for a given instances object.
   a. Input: Instances object.
   b. Output: int array of size 4 in this order [TP, FP, TN, FN].

SMO code examples:

Creating an SVM classifier:

```
SMO smo = new SMO();
PolyKernel kernel = new PolyKernel();
smo.setKernel(kernel);
smo.buildClassifier(data);
```

Setting the degree of a polynomial kernel:

```
PolyKernel kernel = new PolyKernel();
kernel.setExponent(degree);
```

Choosing RBF kernel:

```
RBFKernel kernel = new RBFKernel();
smo.setKernel(kernel);
smo.buildClassifier(data);
```

Setting the gamma parameter of an RBF kernel:

```
RBFKernel kernel = new RBFKernel();
kernel.setGamma(0.3);
```

Classify a new instance:

```
smo.classifyInstance(someInstance);
```

Define the kernel object for the SVM

```
smo.setKernel(kernel);
```

Set the parameter of the slack variable

```
m_smo.setC(c);
```

Your complete output should look like this:

```
For PolyKernel with degree <the degree of the polynomial kernel> the rates are:
TPR = <TRP_Result>
FPR = <FPR_Result>

For RBFKernel with gamma <the gamma value> the rates are:
TPR = <TRP_Result>
FPR = <FPR_Result>

The best kernel is: <Poly or RBF> <kernel parameter (degree or gamma)> <TPR / FPR>

For C <C value> the rates are:
TPR = <TRP_Result>
FPR = <FPR_Result>
```

The yellow part should occur many times, one for each polynomial kernel.

Similarly, the red part should occur for each RBF kernel.

The green part should occur for each C value.

Theoretical questions:

1. Use Lagrange Multipliers to find the maximum and minimum values of the function subject to the given constraints:

   a. $f(x,y) = 2e^{xy}$; constraint: $2x^2 + y^2 = 32$

      Solution:

$$2e^{xy} + \lambda(2x^2 + y^2 - 32)$$

$$\nabla f_x = 2ye^{xy} + 4\lambda x = 0$$

$$\nabla f_y = 2xe^{xy} + 2\lambda y = 0$$

$$\nabla f_\lambda = 2x^2 + y^2 - 32 = 0$$

$$\frac{4\lambda x}{y} = \frac{2\lambda y}{x}$$

$$2x^2 = y^2$$

$$y^2 + y^2 - 32 = 0$$

$$y^2 = 16 \rightarrow y = \pm 4$$

$$x^2 = 8 \rightarrow x = \pm\sqrt{8}$$

$$max = 2e^{4\sqrt{8}} \qquad min = 2e^{-4\sqrt{8}}$$

b. $f(x,y) = x^2 + y^2$; constraint: $y - \cos x = 0$

<u>*Solution*</u>:

$$x^2 + y^2 + \lambda(y - \cos x)$$

$$\nabla f_x = 2x + \lambda \sin x = 0$$

$$\nabla f_y = 2y + \lambda = 0$$

$$\nabla f_\lambda = y - \cos x = 0$$

$$\frac{2x}{\sin x} = 2y \rightarrow y = \frac{x}{\sin x}$$

$$\frac{x}{\sin x} - \cos x = 0$$

$$x = \sin x \cos x$$

$$\sin 2x = 2x$$

$$x = 0, y = 1, \lambda = -2$$

$$min = 1$$

2.

a. Consider two kernels K_1 and K_2, with the mappings φ_1 and φ_2 respectively. Show that $K = K_1 + K_2$ is also a kernel and find its corresponding φ.

<u>*Solution*</u>:

x, y are two vectors in the lower n-dimension.

Lets d_1, d_2 be the higher dimension respectively.

By definition:

$$K_1(x,y) = \varphi_1(x) \cdot \varphi_1(y) = \sum_{i=1}^{d_1} \varphi_1(x)_i \varphi_1(y)_i$$

$$K_2(x,y) = \varphi_2(x) \cdot \varphi_2(y) = \sum_{i=1}^{d_2} \varphi_2(x)_i \varphi_2(y)_i$$

$$K(x,y) = K_1(x,y) + K_2(x,y) = \sum_{i=1}^{d_1} \varphi_1(x)_i \varphi_1(y)_i + \sum_{i=1}^{d_2} \varphi_2(x)_i \varphi_2(y)_i$$

$$= \varphi_1(x)_1 \varphi_1(y)_1 + \cdots + \varphi_1(x)_{d_1} \varphi_1(y)_{d_1} + \varphi_2(x)_1 \varphi_2(y)_1 + \cdots + \varphi_2(x)_{d_2} \varphi_2(y)_{d_2}$$

$$= \varphi(x) \cdot \varphi(y)$$

Where $\varphi(x) = \left(\varphi_1(x)_1, \dots, \varphi_1(x)_{d_1}, \varphi_2(x)_1, \dots, \varphi_2(x)_{d_2}\right)$

Q.E.D

b. Consider a kernel K_1 and its corresponding mapping φ_1 that maps from the lower space R^n to a higher space R^m (m>n). We know that the data in the higher space R^m, is separable by a linear classifier with the weights vector w.

Given a different kernel K_2 and its corresponding mapping φ_2, we create a kernel $K = K_1 + K_2$ as in section a above. Can you find a linear classifier in the higher space to which φ, the mapping corresponding to the kernel K, is mapping?

If YES, find the linear classifier weight vector.

If NO, prove why not.

Solution:

We saw in a that

$$\varphi(x) = \left(\varphi_1(x)_1, \dots, \varphi_1(x)_{d_1}, \varphi_2(x)_{d_1+1}, \dots, \varphi_2(x)_{d_1+d_2}\right) = [\varphi_1(x), \varphi_2(x)]$$

We also know that the linear separator w separates the instances in R^m.

Given the mapping φ above, notice that R^m is a subspace in the space corresponds to φ, we will define the linear classifier to be:

$$w' = (w_1, \dots, w_m, 0, \dots, 0)$$

This linear classier will give the same result as w where we will calculate the dot product with each instance – therefore it will separate the data as w did.

c. What is the dimension of the mapping function φ that corresponds to a polynomial kernel $K(x,y) = (\alpha x \cdot y + \beta)^d$ $(\alpha, \beta \neq 0)$, where the lower dimension is n?

Solution:

The dimension of the mapping φ is: $\binom{n+d}{n}$ as shown in class.

d. Given two polynomial kernels

$K_1(x,y) = (\alpha_1 x \cdot y + \beta_1)^d$ and $K_2(x,y) = (\alpha_2 x \cdot y + \beta_2)^d$ (note that d is the same in both kernels), with the corresponding mappings φ_1 and φ_2. Find a mapping φ that corresponds to the kernel $K = K_1 + K_2$ and that has the same dimension as φ_1 and φ_2.

Solution:

$$(\alpha_1 x \cdot y + \beta_1)^d + (\alpha_2 x \cdot y + \beta_2)^d$$

$$= \sum_{i=0}^{d} \binom{d}{i} {\alpha_1}^i {b_1}^{d-i} (x \cdot y)^i + \sum_{i=0}^{d} \binom{d}{i} {\alpha_2}^i {b_2}^{d-i} (x \cdot y)^i$$

$$= \sum_{i=0}^{d} \binom{d}{i} \left({\alpha_1}^i {b_1}^{d-i} + {\alpha_2}^i {b_2}^{d-i}\right)(x \cdot y)^i$$

$$= \sum_{i=0}^{d} \binom{d}{i} \left({\alpha_1}^i {b_1}^{d-i} + {\alpha_2}^i {b_2}^{d-i} \right) \sum_{k_1+k_2+\cdots+k_n=i} \binom{i}{k_1, k_2, \ldots, k_n} \prod_{t=1}^{n} {x_t}^{k_t} {y_t}^{k_t}$$

$$= \sum_{i=0}^{d} \binom{d}{i} \left({\alpha_1}^i {b_1}^{d-i} + {\alpha_2}^i {b_2}^{d-i} \right) \sum_{k_1+k_2+\cdots+k_n=i} \frac{i!}{k_1!\, k_2! \cdots k_n!} \prod_{t=1}^{n} {x_t}^{k_t} {y_t}^{k_t}$$

$$= \left({b_1}^d + {b_2}^d \right) + \left(d\left({\alpha_1}^1 {b_1}^{d-1} + {\alpha_2}^1 {b_2}^{d-1} \right) \right) (x_1 y_1 + \cdots + x_n y_n)$$

$$+ \binom{d}{2} \left({\alpha_1}^2 {b_1}^{d-2} + {\alpha_2}^2 {b_2}^{d-2} \right) ({x_1}^2 {y_1}^2 + \cdots + {x_n}^2 {y_n}^2 + 2x_1y_1x_2y_2 + \cdots$$

$$+ 2x_1y_1x_ny_n + \cdots + 2x_{n-1}y_{n-1}x_ny_n)$$

$$+ \cdots$$

$$\varphi$$

$$= \Bigg(\sqrt{\left({b_1}^d + {b_2}^d \right)}, \sqrt{\left(d\left({\alpha_1}^1 {b_1}^{d-1} + {\alpha_2}^1 {b_2}^{d-1} \right) \right)}\, x_1, \ldots, \sqrt{\left(d\left({\alpha_1}^1 {b_1}^{d-1} + {\alpha_2}^1 {b_2}^{d-1} \right) \right)}\, x_n,$$

$$\sqrt{\binom{d}{2} \left({\alpha_1}^2 {b_1}^{d-2} + {\alpha_2}^2 {b_2}^{d-2} \right) {x_1}^2}, \ldots, \sqrt{\binom{d}{2} \left({\alpha_1}^2 {b_1}^{d-2} + {\alpha_2}^2 {b_2}^{d-2} \right) {x_n}^2},$$

$$\sqrt{2\binom{d}{2} \left({\alpha_1}^2 {b_1}^{d-2} + {\alpha_2}^2 {b_2}^{d-2} \right) x_1 x_2}, \ldots, \sqrt{2\binom{d}{2} \left({\alpha_1}^2 {b_1}^{d-2} + {\alpha_2}^2 {b_2}^{d-2} \right) x_{n-1} x_n}, \ldots \Bigg)$$

* Each element composes from $\sqrt{\binom{d}{i}\left({\alpha_1}^i {b_1}^{d-i} + {\alpha_2}^i {b_2}^{d-i} \right)}$ (with the relevant i) and the relevant multinomial coefficient

e. Consider the space $S = \{1,2, \ldots N\}$ for some finite N (each instance in the space is a 1-dimension vector and the possible values are 1, 2, …, N) and the function $f(x, y) = \min(x, y)$.

Find the mapping φ such that:

$$\varphi(x) \cdot \varphi(y) = \min(x, y)$$

For example, if the instances are $x = 3, y = 5$, for some $N \geq 5$, then:

$$\varphi(x) \cdot \varphi(y) = \varphi(3) \cdot \varphi(5) = \min(3,5) = 3$$

Solution:

Given x, y and N the dimension of mapping φ will be N and it will be construct as follow:

For any number x (=instance) the mapping will be x times 1 and then $N - x$ times 0. For example:

If $x = 3, y = 5$ and $N = 6$ the mappings will be:

$$\varphi(x) = \varphi(3) = (1,1,1,0,0,0)$$

$$\varphi(y) = \varphi(5) = (1,1,1,1,1,0)$$

$$\varphi(x) \cdot \varphi(y) = \varphi(3) \cdot \varphi(5) = 1 + 1 + 1 + 0 + 0 + 0 = 3 = \min(3,5)$$

3. Find the Kernel function for the following mapping:

   a. $\varphi(x) = (x_1^3, x_2^3, \sqrt{3}x_1^2x_2, \sqrt{3}x_1x_2^2, 3x_1^2, 3x_2^2, 3\sqrt{2}x_1x_2, 3\sqrt{3}x_1, 3\sqrt{3}x_2, 3\sqrt{3})$

   Solution:

$$\varphi(x) \cdot \varphi(y) = x_1^3y_1^3 + x_2^3y_2^3 + 3x_1^2x_2y_1^2y_2 + 3x_1x_2^2y_1y_2^2 + 9x_1^2y_1^2 + 9x_2^2y_2^2$$
$$+ 18x_1x_2y_1y_2 + 27x_1y_1 + 27x_2y_2 + 27$$
$$= (x \cdot y)^3 + 9(x \cdot y)^2 + 27x \cdot y + 27 = (x \cdot y + 3)^3 = K(x, y)$$

   b. $\varphi(x) = (\sqrt{5}x_1^2, \sqrt{5}x_2^2, \sqrt{10}x_1x_2, \sqrt{8}x_1, \sqrt{8}x_2, \sqrt{5})$

   Solution:

$$\varphi(x) \cdot \varphi(y) = 5x_1^2y_1^2 + 5x_2^2y_2^2 + 10x_1x_2y_1y_2 + 8x_1y_1 + 8x_2y_2 + 5$$
$$= 5(x_1^2y_1^2 + 2x_1x_2y_1y_2 + x_2^2y_2^2) + 8x_1y_1 + 8x_2y_2 + 5$$
$$= 5(x_1y_1 + x_2y_2)^2 + 8x_1y_1 + 8x_2y_2 + 5$$
$$= (x_1y_1 + x_2y_2)^2 + 4x_1y_1 + 4x_2y_2 + 4 + 4(x_1y_1 + x_2y_2)^2 + 4x_1y_1 + 4x_2y_2 + 1$$
$$= (x \cdot y)^2 + 4x \cdot y + 4 + 4(x \cdot y)^2 + 4x \cdot y + 1$$
$$(x \cdot y + 2)^2 + (2x \cdot y + 1)^2 = K_1(x, y) + K_2(x, y) = K(x, y)$$

You should hand in a SVM.java, MainHW5.java, the hw5 excel file filled with the relevant results and this word document with answers to the theoretical questions. All of these files should be placed in a hw_5_##id##.zip folder with your id.