Assignment 6

Machine Learning, Summer term 2014, Ulrike von Luxburg

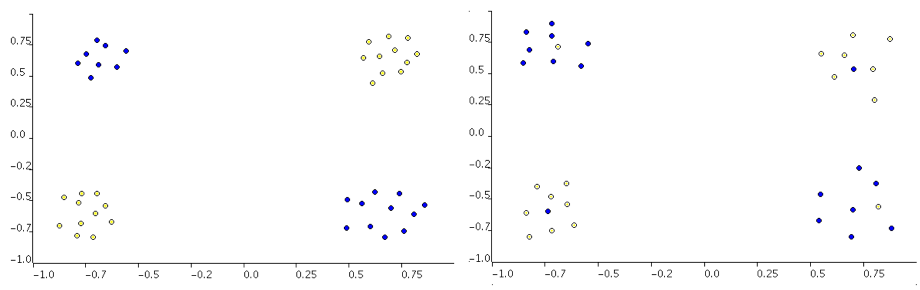
To be discussed in exercise groups on May 26-28

Exercise 1 (Play with SVM, 2+2 points) In this exercise, you would play with a Java implementation of SVM, which is available as an applet:

`www.ml.inf.ethz.ch/education/lectures_and_seminars/annex_estat/Classifier/JSupportVectorApplet.html`

Note: In the new version of Java, you may encounter a security alarm from your browser. Add `http://www.ml.inf.ethz.ch/` to the list of your trusted websites. For more information, see
`http://www.java.com/en/download/faq/exception_sitelist.xml`
`http://www.java.com/en/download/help/jcp_security.xml`

Set the training points as depicted in Figure 1-a.



(a) (b)

Figure 1: (a) Train data. (b) Train data with outlier.

(a) Train the SVM with the following settings. Capture the output screen for your report.
   - Linear kernel (Simple Dot Product) with $C = 100$.
   - Polynomial kernel of degree 2, and degree 8. Choose a proper C.
   - Gaussian kernel (Radial Basis Function): In the applet, they use a different notation $\beta = 1/(2\sigma^2)$. Try $\beta = 0.01, 1, 10, 100$. Choose a proper C.

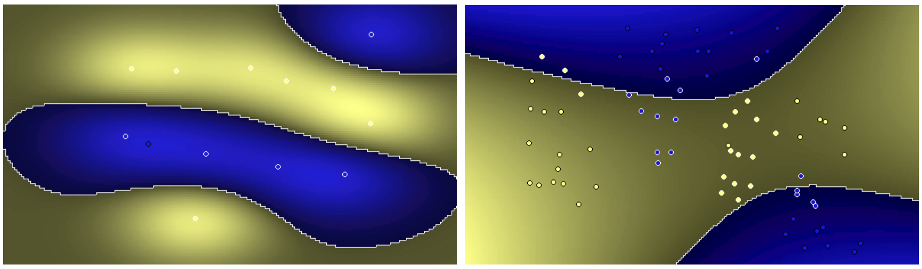
(b) Add noise to your training data as depicted in Figure 1-b. Try the Gaussian kernel with $\beta = 10$ and $C = 0, 10, 1000$. Based on your observation, describe the effect of the parameter C.

Exercise 2 (Understanding kernel SVM, 1+1+1 points) The output of kernel SVM in two problems with different parameters and kernels are depicted in Figure 2-a and 2-b. For each figure, answer the following questions:

(a) Which type of kernel is used: linear, polynomial or rbf?

(b) Argue if this is a good classifier? How should we change the parameters of the classifier to avoid this problem?

(c) Guess the support vectors in Figure 2-a.



(a) (b)

Figure 2

Exercise 3 (SVM in matrix form, 1+1 points) Write the primal and the dual SVM in a matrix notation. Use the symbol $\mathbf{1}_d$ to show the vector $[1, 1, ..., 1]^T$ with length d. Note that for vectors $a, b \in \mathbb{R}^d$ and matrix X we have

$$a^T b = \sum_i a_i b_i \;\; ; \;\; a^T X b = \sum_{i,j} a_i b_j X_{i,j}$$
$$(XX^T)_{i,j} = \langle X_{i,*}, X_{*,j} \rangle.$$

(a) **Soft margin linear SVM: Primal** **Dual**

$$\min_{\mathbf{w},\xi,b} \quad \tfrac{1}{2} \cdots\cdots + \tfrac{C}{n} \cdots\cdots \mathbf{1}_n$$
$$y_i(\mathbf{w}^{\mathrm{T}}\mathbf{x_i} + b) \geq 1 - \xi_i \quad i = 1, \ldots, n$$
$$\xi_i \geq 0 \quad i = 1, \ldots, n$$

$$\max_{\alpha} \quad \cdots\cdots - \tfrac{1}{2}\alpha^{\mathrm{T}} .... \alpha$$
$$\cdots\cdots = 0$$
$$0 \leq \alpha_i \leq \tfrac{C}{n} \quad i = 1, \ldots, n$$

(b) **Kernel SVM: Primal** **Dual**

$$\min_{\mathbf{w},\xi,b} \quad \tfrac{1}{2} \cdots\cdots + \tfrac{C}{n} \cdots\cdots$$
$$y_i(\mathbf{w}^{\mathrm{T}}\mathbf{\Phi}(\mathbf{x_i}) + b) \geq 1 - \xi_i \quad i = 1, \ldots, n$$
$$\xi_i \geq 0 \quad i = 1, \ldots, n$$

$$\max_{\alpha} \quad \cdots\cdots - \tfrac{1}{2}\alpha^{\mathrm{T}} .... \alpha$$
$$\cdots\cdots = 0$$
$$0 \leq \alpha_i \leq \tfrac{C}{n} \quad i = 1, \ldots, n$$

Exercise 4 (Building new kernels, 0.5+0.5+1+0+1 points)
Assume that $K_1, K_2 : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ are kernel functions. Which of the following functions are also a valid kernel? Prove or bring a counterexample.

(a) $K = \alpha K_1$ for $\alpha > 0$

(b) $K = K_1 + K_2$

(c) $K = K_1 - K_2$

(d) $K(x, y) = K_1(x, y) \cdot K_2(x, y)$ (optional)

(e) $K(x, y) = f(x)K_1(x, y)f(y)$ for any function $f : \mathcal{X} \to \mathbb{R}$.

Exercise 5 (Polynomial kernel, 1.5+1.5 points) Consider the second degree polynomial kernel function $K(x, y) = (x^T y + 1)^2$ with inputs $x, y \in \mathbb{R}^2$.

(a) Show that the corresponding feature map function is $\Phi(x) = (1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, x_2^2, \sqrt{2}x_1x_2)^T$ where $x = (x_1, x_2)^T \in \mathbb{R}^2$.

(b) If we use the second degree polynomial kernel for inputs from $\mathbb{R}^d$, what would be the dimensionality of the corresponding feature space?

Exercise 6 (MATLAB experiment, 1+1+1+1+1 points) In optimization with CVX, you can also access the dual variables. For example in soft margin SVM

```
cvx_begin
    variables w(d) b xi(n)
    dual variable lambda
    minimize 1/2*sum(w.*w) + C/n*sum(xi)
    lambda : Y.*(X*w + b) >= 1 - xi;
    xi >= 0;
cvx_end
```

(a) Use the following training data and set $C = 1$. Find the optimal primal w^* and the optimal dual variables λ^*.

```
X = [-3 3;-3 2;-2 3;-1 1;1 3;2 2;2 3;3 1];
Y = [-1 -1 -1 -1 1 1 1 1]';
```

(b) From dual variable λ, find the support vectors (Support vectors are training points which the constraint is active on them: λ_i is larger than zero). Note that matlab is a numerical package, so in this example you can count values smaller than 10^{-6} as zero.

(c) Check the KKT condition. To do this, you need to check that $\lambda_i^* \left(Y_i(w^{*T}X_i + b^*) - 1 + \xi_i^*\right)$ is zero ($< 10^{-6}$) for all i.

Here you have a CVX implementation of dual SVM with linear kernel $K(x, y) = x^T y$

```
K = X*X';
cvx_begin
    variables alpha(n) %you don't have anything with size d
    maximize( sum(alpha) -  0.5*quad_form(Y.*alpha,K) )
       alpha>=0;
       alpha<=C/n;
       alpha'*Y==0;
cvx_end
```

(d) Verify that variables `lambda` and `alpha` are approximately equal.

(e) Reconstruct the primal variables w and b from `alpha,X,Y`. Is the result the same as the one you got from the primal?