



University of Tehran  
School of Electrical and Computer Engineering



# Pattern Recognition

## Assignment 5

Due Date: Dey 20, 1398

Corresponding TAs:

Melika Kheirieh - [melikakheirieh@gmail.com](mailto:melikakheirieh@gmail.com)

Arman Panahi - [armanp30@gmail.com](mailto:armanp30@gmail.com)

Ala Emrani – [emrani.ala@ut.ac.ir](mailto:emrani.ala@ut.ac.ir)

Dey 98

## PROBLEM 1

Show that for the case of two linearly separable classes the hyperplane obtained as the SVM solution is the same as that bisecting the segment joining two closest points between the convex hulls of the classes.

## PROBLEM 2

Suppose  $k_1$  and  $k_2$  are valid (symmetric, positive definite) kernels on  $X$ . Then, show the following are valid or invalid kernels:

- A)  $K(u, v) = \alpha k_1(u, v) - \beta k_2(u, v)$ , for  $\alpha, \beta \geq 0$
- B)  $K(u, v) = k_1(u, v) k_2(u, v)$
- C)  $K(u, v) = \exp\left(\frac{-\|u-v\|^2}{\sigma^2}\right)$
- D)  $K(u, v) = g(u) g(v)$ , for  $g: x \rightarrow \mathbb{R}$
- E)  $K(u, v) = f(k_1(u, v))$ , where  $f$  is a polynomial with positive coefficients.

## PROBLEM 3

In condition that no misclassification is allowed, SVM problem form would be as follows (Hard Margin SVM):

$$\min \frac{\|w\|^2}{2}$$
$$s.t.: (x^i, y^i): y^i(w^T x^i + w_0) \geq 1$$

1. In case complete classification is not possible, make condition which points from each class is allowed to cross the borders. Rewrite new form of problem (Soft Margin SVM) with variable  $\xi_i$  and condition  $\xi_i \geq 0$ .
2. Explain what is the meaning of each case ( $\xi_i = 0$ ,  $\xi_i = 1$ ,  $0 < \xi_i < 1$ ,  $\xi_i > 1$ ) in the previous section.

## PROBLEM 4

show that if the soft margin SVM cost function is chosen to be

$$\frac{1}{2} \|w\|^2 + \frac{C}{2} \sum_{i=1}^N \xi_i^2$$

the task can be transformed into an instance of the class-separable case problem

## PROBLEM 5

You are provided with a dataset of 1000 customers of a retail store in new York. (retailMarketing.csv) Description of the fields of this dataset is below:

**AmountSpent:** amount of money that the customer spent in this store.

**Catalogs:** the category of the item in which the customer has been buy.

**Children:** number of children that this customer have had.

**Gender:** gender of the customer.

**History:** this field regards to the activity of the customer.

**Location:** distance of the accommodation of the customer to this store.

**Married:** Marriage of the customer

**OwnHome:** Customer's home is for himself or is rented?

**Salary:** annual salary of the customer

**age:** Middle or young or old?

1. Design a SVM classifier that determine age of the customer according to his/her information. (split data into train and test properly)
2. Use linear SVM with appropriate value of the parameter “c”. Compare the result of this part with part 1.
3. Consider the effect of normalization the data, on part 1 and 2.
4. Test the effect of parameter tuning for RBF-kernel SVM. You need to tune the values of the parameters C and  $\gamma$  using GridSearchCV. What is the best accuracy you get? Draw a diagram showing the CCR versus parameters.

## NOTES

1. Please make sure you reach the deadline because there would be no extra time available.
2. Late policy would be as bellow:
  - Every student has a budget for late submission during the semester. This budget is two weeks for all the assignments.
  - Late submission more than two weeks may cause lost in your scores.
3. Analytical problems can be solved on papers and there is no need to type the answers. The only thing matters is the quality of your pictures. Scanning your answer sheets is recommended. If you are using your smartphones you may use scanner apps such as CamScanner or google drive application.
4. Simulation problems need report as well as source codes and results. This report must be prepared as a standard scientific report.
5. You have to prepare your final report including the analytical problems answer sheets and your simulation report in a single pdf file.
6. Finalized report and your source codes must be uploaded to the course page as a “.zip” file (not “.rar”) with the file name format as bellow:  
**PR\_Assignment #[Assignment Number]\_Surname\_Name\_StudentID.zip**
7. Plagiarisms would be strictly penalized.
8. You may ask your questions from corresponding TAs.