THE CHINESE UNIVERSITY
OF HONG KONG (SHENZHEN)
School of Science and Engineering

Tutorial - 10

Fundamental of Artificial Intelligence **CSC3180**

1. Answer the following questions:

(a) What is difference between clustering and classification? (unsupervised and supervised)

Classification is the process of classifying the data with the help of class labels (supervised signals). On the other hand, Clustering is similar to classification but there are no predefined class labels. Classification is geared with supervised learning. As against, clustering is also known as unsupervised learning.

BASIS FOR COMPARISON	CLASSIFICATION	CLUSTERING
Basic	This model function classifies the data into one of numerous already defined definite classes.	This function maps the data into one of the multiple clusters where the arrangement of data items is relies on the similarities between them.
Involved in	Supervised learning	Unsupervised learning
Training label	Provided	Not provided

(b) What is the advantage and disadvantage of the k-nearest neighbors algorithm? (L11: P13)

Advantages:

- does not need any prior knowledge about the shape of the decision function
- simple to implement (**not obvious learning process**), strong asymptotic guarantees
- performs well with large training sets & low dimensional data

Disadvantages:

- computationally costly for prediction
- needs a large number of examples to adapt to the true decision boundary
- the choice of the distance, i.e., *k*, is critical

(c) What is the influence of the value of K in k-nearest neighbors algorithm? (L11: P11-12)

k plays the role of a capacity control parameter:

- k =1 (too small) high sensitivity to noise & outliers.

- Larger k produces smoother boundary effect & can reduce the impact of class label noise.
- But when k is too large, we always predict the majority class.

(d) What is the difference between the k-nearest neighbor and Support Vector Machine? (no learning and learning)

1. K-Nearest Neighbor (KNN) is the simplest supervised learning method (no learning process)
2. Support Vector Machine (SVM) maximizes the margin width between different classes and introduce a slack variable to make model realistic if samples from different classes are not linearly classifiable.

(e) What is sparse? How about its advantages? (L11: P32)

1. Sparse representation-based classification (SRC) method looks for the sparsest representation of a test example in a dictionary composed of all training data across all classes.
2. Sparse representation-based methods have been shown to be effective for classification

Advantages:

- Robust to illumination and occlusion
- With the highest accuracy comparing with traditional methods, such as SVM, KNN

Disadvantages:

- It is time-consuming to search the optimal solution

2. There are three categories and in each category there are three points, shown as follows. Give two point x = [0, 0, 4, 2] and y = [-1, 4, 1, 5], please judge these two points belong to which class. (Euclidean Distance)(K=1)

a_1 = [1, 0, 3, 1]; a_2 = [1, 2, 5, 0]; a_3 = [0, 1, 3, 1]

b_1 = [5, -1, 3, 3]; b_2 = [4, 0, 5, 5]; b_3 = [6, 1, 4, 3]

c_1 = [-1, 4, 1, 3]; c_2 = [0, 3, 1, 4]; c_3 = [-1, 3, 2, 3]

Solution:

$D(x,a_1) = 1.7321$, $D(x,a_2) = 3.1623$, $D(x,a_3) = 1.7321$

$D(x,b_1) = 5.2915$, $D(x,b_2) = 5.0990$, $D(x,b_3) = 6.1644$

$D(x,c_1) = 5.1962$, $D(x,c_2) = 4.6904$, $D(x,c_3) = 3.8730$

Thus x belongs to the first class.

Similarly,

$D(y,a_1) = 6.3246$, $D(y,a_2) = 7.0000$, $D(y,a_3) = 5.4772$

$D(y,b_1) = 8.3066$, $D(y,b_2) = 7.5498$, $D(y,b_3) = 8.4261$

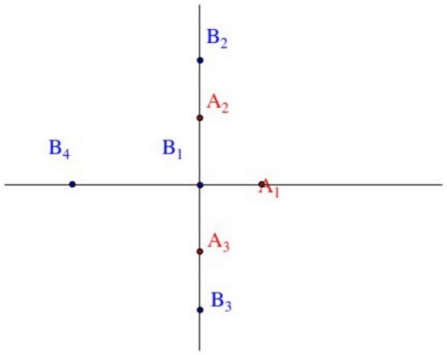
$D(y,c_1) = 2.0000$, $D(y,c_2) = 1.7321$, $D(y,c_3) = 2.4495$

Thus y belongs to the third class.

3. **Assume that in a 2-D space, there are two categories, A and B. For A category, there are three training points (red). For B category, there are four training points (blue)**

   (1) According to KNN algorithm, how many classification planes among these two categories?

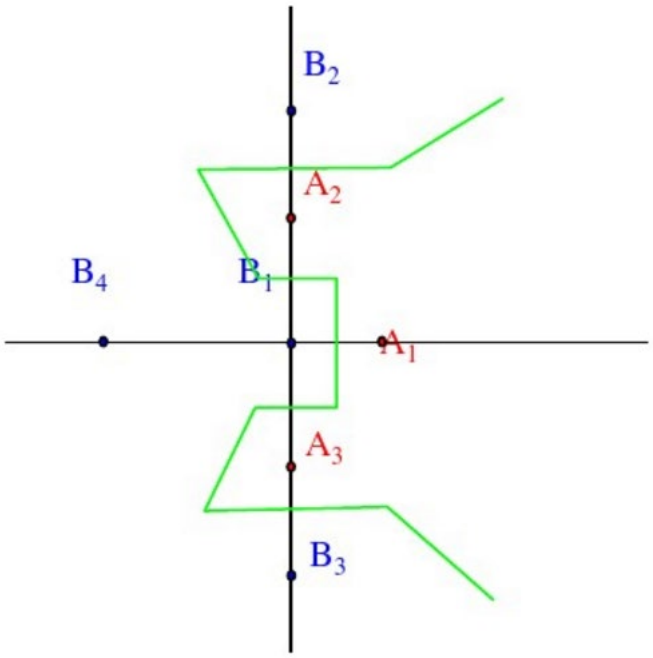
   (2) Please draw classification planes in the picture.



Solution:

For each two training points which belong to different categories, they can construct a nearest classifier if they are the neighbors for the testing point. And the perpendicular line between them would be the classification plane. Thus the max number of the classification planes constructed by the points in A and B categories are 3×4=12.

One example for classifying A and B is shown as follows, and there are 9 planes.



4. **Please describe the main steps of Sparse Representation Classification.**

Solution:

Step1: Give the parameter λ, the training and testing samples X and y, where $X \in R^{d\times n}$ and $y \in R^{d\times 1}$ (d is the dimension of the data and n is the number of the training data).

Step2: Minimize the following formulation

$$\|y - X\alpha\|_2^2 + \lambda\|\alpha\|_1$$

Then we can get the representation coefficient $\hat{\alpha}$ by optimization algorithm.

Step3: Compute the residual for each class

$$r_k = \|y - X_k\hat{\alpha}_k\|_2$$

Step4: Output the identity of y.

$$\text{Identity}(y) = \min_k\{r_k\}$$

5. **Using an example of face recognition, please describe how Sparse Representation Classification works.**

Solution:

In SRC, if a testing sample belongs to a certain category, it can be well represented by some samples belonging to this category in the training data. Due to the L_1 norm, the representation coefficients would be sparse. Thus, many elements in the representation coefficients would be zeros and only a few of elements are non-zero. Furthermore, if the testing sample belongs to the t-th category, most of these non-zeros coefficients are associated with the training samples belonging to the t-th category. According to the classification rule, $r_t(y) = \|A_t\hat{\alpha}_t - y\|$ would be the smallest. Thus, we can get the predicted label for the testing sample y.