Machine Learning & Machine Learning (extended)

(Practice Exercise Sheet 2)

Please put in some effort to solve these exercises to get the most out of topics studied and to increase your chance of doing well on the Canvas Quiz test. This document is only an example of the kind of questions you may be asked in the Quiz, but it's no way an exhaustive list of all kind of questions. Some of these questions will be solved in the class on Thursday as preparation before the Quiz.

Clustering

Question: Use K-means algorithm and squared Euclidean distance measure to cluster the following 2-dimensional objects in to 3 clusters. The points are described as: O1 (2,10), O2 (2,5), O3 (8,4), O4 (5,8), O5 (7,5), O6 (6,4), O7 (1,2), O8 (4,9). Suppose that initial cluster centres are O1, O4, and O7. Draw a 10x10 grid with all the 8 objects with initialized cluster centres marked.

a) Run the K-means algorithm steps for 1 iteration and show: (i) the clusters (i.e. objects belonging to each cluster), (ii) the centres of new clusters, and (iii) draw a 10x10 grid with all the 8 objects and show the clusters and centres after the first iteration.
b) Using graphical drawing, illustrate the algorithm iterations until the algorithm converges. How many iterations it takes to converge?

Question: Repeat the above with Manhattan distance measure.

Question: Use min/single link to perform agglomerative clustering by showing the dendrogram for the data described by the distance matrix below:

	A	B	C	D
A	0	1	4	5
B		0	2	6
C			0	3
D				0

Note: the height of each "junction" in the dendrogram represents the distance between the pair of clusters.

Question: Repeat the above with max/complete link to perform agglomerative clustering.

Question: Use single link agglomerative clustering to cluster the following 8 objects by showing the dendrograms: O1 (2,10), O2 (2,5), O3 (8,4), O4 (5,8), O5 (7,5), O6 (6,4), O7 (1,2), O8 (4,9).

Question: Use complete link agglomerative clustering to cluster the following 8 objects by showing the dendrograms: O1 (2,10), O2 (2,5), O3 (8,4), O4 (5,8), O5 (7,5), O6 (6,4), O7 (1,2), O8 (4,9).

Question: What is the goal of clustering, and how it differs from classification?

Question: State three application areas of k-means clustering.

Question: Describe in what situation the conventional k-means algorithm would fail to cluster the data. Can you suggest a modification to overcome the problem?

Question: Suppose you have run k-means clustering on an available data set. Later you get more data points which are observed over similar attributes/features. Can we cluster the new data points using the results of first run of k-means algorithm?

Decision Boundary as a Straight Line

Question: Given the equation of a straight line $y = -0.75x + 4$: (i) draw this line, (ii) represent it in the form $\boldsymbol{w}^T\boldsymbol{x} + b = 0$ by finding the values of $\boldsymbol{w}$ and b, (iii) draw parallel line $\boldsymbol{w}^T\boldsymbol{x} + b = 1$, (iv) draw parallel line $\boldsymbol{w}^T\boldsymbol{x} + b = -1$, (v) show the mathematical working and graphical visualization of a couple of example points for which $\boldsymbol{w}^T\boldsymbol{x} + b > 0$, (vi) show the mathematical working and graphical visualization of a couple of example points for which $\boldsymbol{w}^T\boldsymbol{x} + b < 0$.

Support Vector Machines

Question: The following 2-dimensional data is to be used for training an SVM:

Class one: (1,1), (2,2), (2,0)

Class two: (0,0), (1,0), (0,1)

(i) Plot the training points and, by visual inspection, determine the position of the optimal margin decision boundary.

(ii) List the support vectors.

Question: Consider training data of 1-dimensional points from two classes:

Class 1: -5,5

Class 2: -2,1

(i) Are the two classes linearly separable?

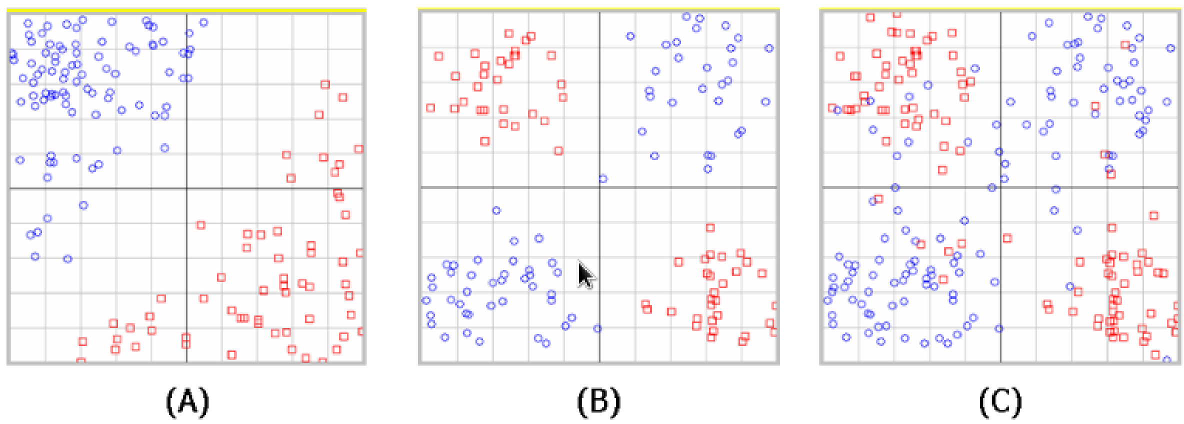
(ii) Consider the transformation $\varphi: R \rightarrow R^2$, $\varphi(x) = (x, x^2)$. Transform the data and plot these transformed points. Are these linearly separable?

(iii) Draw the optimal separating hyper-plane in the transformed space, and explain in one or two sentences how this linear boundary helps us to separate the original data points.

Question: What is the main idea behind a linear SVM? Illustrate your explanation by drawing a figure.

Question: How can a linear SVM work for learning to classify data that is not linearly separable?

Question: Consider the three data sets illustrated below:



Each point has two numeric features (i.e. the x and y coordinates of the points). The circles and the squares represent two different classes.

(i) If you were to use SVM on these data sets, which data set will require a non-linear kernel to be used?

(ii) If you were to use SVM with Gaussian kernel on these data sets, how would you set the width parameter of the kernel?

Question: The XOR problem is to learn the function from the following input points to their class labels:

Class : (1,1), (-1,-1)

Class 2: (1,-1), (-1,1)

We know that support vector machine (SVM) with a kernel can solve this problem by mapping the points in a higher dimensional space. But higher dimensional spaces are difficult to visualise, and we would like to construct a support vector machine that classifies these points correctly in a 2-dimensional input space. Is this possible? If so, how? If not, explain why not.

Hint: Try to come up with a feature-transformation that stays in 2D (i.e. maps R^2 into R^2) and makes the classes linearly separable.

Classification

Question: Your employer wants to classify properties to decide which one to buy. They have a lot of training data, and each property is described compactly by a handful of attributes. Can you recommend two classification methods to try? Provide reasoning for your choices.

Question: Your employer wants to classify job applications to decide which applicants to invite for interview, using a large number of features that describe the application. So far they have just a few applications that can be used as training data, though. Can you recommend two classification methods to try? Provide reasoning for your choices.

Covariance and Bayesian Classification

Question: Let's consider the mean and covariance of two class dataset below.

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \boldsymbol{\Sigma}_1 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \boldsymbol{\mu}_2 = \begin{bmatrix} 8 \\ 8 \end{bmatrix}, \boldsymbol{\Sigma}_1 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

Can you make a rough drawing of how the point cloud of each class would be modelled with multivariate Gaussian pdf? What will be the shape of the decision boundary with a Gaussian classifier, in the cases of with and without naïve assumption?

Question: Repeat the above with the following mean and covariance of two class dataset.

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \boldsymbol{\Sigma}_1 = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}, \boldsymbol{\mu}_2 = \begin{bmatrix} 8 \\ 8 \end{bmatrix}, \boldsymbol{\Sigma}_1 = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

Question: Repeat the above with the following mean and covariance of two class dataset.

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \boldsymbol{\Sigma}_1 = \begin{bmatrix} 2 & 2 \\ 2 & 3 \end{bmatrix}, \boldsymbol{\mu}_2 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \boldsymbol{\Sigma}_1 = \begin{bmatrix} 2 & -2 \\ -2 & 4 \end{bmatrix}$$

Question: Consider a 2-class problem. Let $p(\boldsymbol{x}|c)$, where $c \in (1,2)$, be a multivariate Gaussian pdf, with mean vector $\boldsymbol{\mu}_c$, and covariance matrix $\boldsymbol{\Sigma}_c$. Let us assume that the covariance matrix is same for both classes with same variance for all attributes and the non-diagonal entries are zero. For this case, show with a graphical illustration the modelled shape of point cloud of each class and the shape of the decision boundary of the Gaussian classifier.

Question: Is it possible for a Gaussian classifier to implement a non-linear decision boundary? If so, draw an example. If not, explain why not.

Question: Is it possible for a Gaussian Naïve Bayes classifier to implement a non-linear decision boundary? If so, draw an example. If not, explain why not.

(Bonus) Question: [Optional] Repeat the above to mathematically show whether it has the form of a linear boundary $\boldsymbol{w}^T\boldsymbol{x} + b$ with some weight vector $\boldsymbol{w}$ and some scalar b.