

Exam for course 3

Surname :

Firstname :

Please report your answers on this page only. Questions are on the following pages.

Question 1 (G-3.1) : Clustering

Question 2 (G-3.2) : Sparse Dictionary Learning

Question 3 (G-3.3) : Applications of unsupervised learning

Question 4 (B-3.1) : Unsupervised learning

Question 5 (B-3.2) : Challenges of Unsupervised learning

Question 6 (B-3.3) : Exploiting unsupervised learning for classification

[Green] Question 1 : Clustering

Report the numbers corresponding to clustering :

1. Clustering can be used to find a decomposition of a huge dataset into a sum of vectors
2. Clusters are necessarily vectors from the original dataset
3. In a distance-based approach, the best clustering for $K = 1$ is obtained by considering the average vector over the whole dataset
4. A dataset can be optimally compressed by considering as many clusters as examples

[Green] Question 2 : Sparse Dictionary Learning

Report the numbers corresponding to sparse dictionary learning (DL) :

1. DL can be used to compress a dataset
2. In dimension d , an optimal dictionary of $K \leq d$ atoms can be easily found by considering a completion of a basis
3. DL can be used to find a decomposition of a huge dataset into a sum of vectors
4. A dataset can be optimally compressed by considering as many dictionary atoms as examples

[Green] Question 3 : Applications of unsupervised learning

Report the numbers corresponding to applications of unsupervised learning :

1. Compression in hard drives
2. Speaker recognition
3. Extracting representative patterns of fake news propagation
4. Finding biomarkers for a neurological disease using many labeled examples of healthy people and patients

[Blue] Question 4 : Unsupervised learning

Report the numbers corresponding to unsupervised learning problems :

1. Finding similarities between vectors
2. Summarizing a huge set of images by finding some representative examples
3. Consider a huge dataset as a matrix and decompose it using an orthogonal basis
4. Compressing a images dataset by grouping them using labels to learn a decomposition

[Blue] Question 5 : Challenges of Unsupervised learning

Report the numbers that almost surely correspond to difficult (=impossible) unsupervised learning problems :

1. Retrieving representative examples of an unknown number of image categories by clustering them without using labels
2. Learn centroids in dimension 20 with 40 examples
3. Finding the best number of dictionary atoms in a dataset with high variance
4. Using euclidean distance to consider centroids based on baricenters in dimension 4 with 2000 examples

[Blue] Question 6 : Exploiting unsupervised learning for classification

Explain in a few sentences how clustering could be used as a first step for classification.