Pattern Recognition and Machine Learning

Home Work 5

Faculty of New Science & Technologies
University Of Tehran
Fall 2019

Part 1

General Homework Policies

1. Due date of this homework is on *Tuesday 14 Dey 98 (4 Jan. 2020)*, so you need to submit it before the due date[midnight 14 Dey] otherwise you won't get any score!
2. Try to budget your time because due dates are hardly changeable, and we will not accept late homework for any reason.
3. You are welcome to collaborate, cooperate, and consult with your classmates provided that you write-up the solutions independently.
4. Don't plagiarize! Write everything in your own words, and properly cite every outside source you use. Taking credit for work as well as ideas that are not your own is plagiarism. Students who plagiarize will not get any score and they will be introduced in the class.
5. Please create reference for all sources(books, papers, websites) which you use.
6. Please create a cover letter for your report which is simply is the Homework#, title of the course, your name, surname, and student number.
7. You may post questions asking for clarifications and alternate perspectives on concepts on piazza or in the class.
8. Email your final file of assignment to *mailto: Sajjadaghapour@ut.ac.ir*, *mailto: e.sadeqi.n@gmail.com* and *ah.havvaei@ut.ac.ir* with subject [PRML hw# Surname] which # indicates number of the home work.

Part 2

Questions

1 Online learning

The unbiased estimates for the covariance of a d-dimensional Gaussian based on n samples is given by,

$$\widehat{\Sigma} = C_n = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i - \mathbf{m}_n)(\mathbf{x}_i - \mathbf{m}_n)^T \tag{2.1}$$

Where $\mathbf{m}_n = \sum_{i=1}^{n}\mathbf{x}_i$. It is clear that it takes $O(nd^2)$ time to compute C_n. If the data points arrive one at a time, it is more efficient to incrementally update these estimates than to recompute from scratch.

(i) Show that the mean vector $\mathbf{m}_{n+1}$ and covariance matrix C_{n+1} can be sequentially updated as follows,

$$\mathbf{m}_{n+1} = \mathbf{m}_n + \frac{1}{n+1}(\mathbf{x}_{n+1} - \mathbf{m}_n) \tag{2.2}$$

$$C_{n+1} = \frac{n-1}{n}C_n + \frac{1}{n+1}(\mathbf{x}_{n+1} - \mathbf{m}_n)(\mathbf{x}_{n+1} - \mathbf{m}_n)^T \tag{2.3}$$

(ii) How much time does the covariance matrix take per sequential update? (Use big-O notation.)

(iii) Show that we can sequentially update the precision matrix(inverse of covariance matrix) using,

$$C_{n+1}^{-1} = \frac{n}{n-1}\left[C_n^{-1} - \frac{C_n^{-1}(\mathbf{x}_{n+1} - \mathbf{m}_n)(\mathbf{x}_{n+1} - \mathbf{m}_n)^T C_n^{-1}}{\frac{n^2-1}{n} + (\mathbf{x}_{n+1} - \mathbf{m}_n)^T C_n^{-1}(\mathbf{x}_{n+1} - \mathbf{m}_n)}\right] \tag{2.4}$$

Hint: notice that the update to C_{n+1} consists of adding a rank-one matrix, namely $\mathbf{u}\mathbf{u}^T$, where $\mathbf{u} = \mathbf{x}_{n+1} - \mathbf{m}_n$. Use the matrix inversion lemma for rank-one updates:

$$(\mathbf{E} + \mathbf{u}\mathbf{v}^T)^{-1} = \mathbf{E}^{-1} - \frac{\mathbf{E}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{E}^{-1}}{1 + \mathbf{v}^T\mathbf{E}^{-1}\mathbf{u}} \tag{2.5}$$

(iv) What is the time complexity per update?

2 Programming Gaussian Classifiers

(i) Write a computer program to perform a linear discriminant analysis (LDA) by fitting a separate Gaussian model per class. Try it out on the IRIS data, and compute the miss-classification error for the test data. The data can be found in following address,
IRIS data set

(ii) Add a pre-processing step. Implement feature reduction using PCA. Then apply LDA on the reduced 2-dimentional data.

(iii) Add a pre-processing step. Implement feature reduction using FDA. Then apply LDA on the reduced 2-dimentional data.

(iv) Compare and analyze results of part i to iii.

(v) Repeat part ii and iii on the vowel dataset given below. Compare these two pre-processors on this dataset.
Vowel test data
Vowel training data
Vowel dataset information

Good Luck!