University of TEHRAN

Machine Learning

Problem Set 9

Hesam Montazeri
Fereshteh Fallah
Mozhgan Mozaffari Legha

Khordad 8, 1399
(May 28, 2020)

Problem 1: Review part
Write your reviews for the whiteboard notes and the slides of the lectures of this week. Write down all formulas and explain in detail each step of the derivations, if applicable.

Problem 2: Conceptual questions
[ISL] chapter 5: questions 2, 6 (programming)

Problem 3: Common statistical hypothesis tests (Optional)
For each of the following tests: one-sample t-test; two-sample t-test, paired t-test, Wilcoxon signed-rank test, Mann-Whitney-Wilcoxon, Pearson's chi-squared test, Fisher's Exact Test, Analysis of variance (the F-test), hypergeometric test, Pearson correlation and Spearman correlation test

(a) Explain the aim of the test in your own words.

(b) Define the null and alternative hypotheses.

(c) Explain the main assumptions of the test.

(d) Write down the test statistic and null distribution.

(e) Give an example application of the test in bioinformatics (you may give the same example for several tests, if applicable).

(f) How do you perform this test using a built-in function in the programming language of your choice?

Important: In addition, draw your own flowchart on how to choose a statistical test from the above list.

Problem 4: Programming: COVID-19 Diagnosis Based on CT Scans
Your task is to assess the performance of SVM with two different kernels, ridge regression and LASSO for making diagnosis based on the CT images publicly available (here) [1] using two different feature extraction methods namely *scale-invariant feature transform* (SIFT) [2] and *Speeded-Up Robust Features* (SURF) [3]. You may need to perform some preprocessings on raw images before feature extraction and classification.

We encourage discussing the problems with other students, however, similarity between solutions is not allowed. (**Important**) Studying any online or previous solutions, no matter to what extent, is

strictly forbidden and is considered as a violation of the academic honor code. Submit your solutions by Khordad 13, 1399.

References

[1] Jinyu Zhao, Yichen Zhang, Xuehai He, and Pengtao Xie. Covid-ct-dataset: a ct scan dataset about covid-19. *arXiv preprint arXiv:2003.13865*, 2020.

[2] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[3] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *European conference on computer vision*, pages 404–417. Springer, 2006.