

Hand-on Exercise 4

Guidelines

- (1) Each group consists of at most two students. Members in the same team will receive the same grade. (Please give the names of your group members to the TA.)
- (2) Dataset “data-ex4-s25.csv” is posted on the course page.
- (3) Each team may use MATLAB or Python as the development platform.
- (4) The performance will be judged by the quality of solutions obtained, the associated analysis, and discussions.
- (5) The deadline for turning in your report is April 29 (Tuesday) at 3:00 PM.

Assignment Description

Background: A hospital aims to predict which patients admitted for chest pain are at high risk of requiring expensive cardiac interventions (e.g., cardiac catheterization, stenting, or bypass surgery) within 30 days of admission. The average cost of such an intervention is \$20,000. The hospital estimates that the cost of not identifying a high-risk patient and delaying necessary treatment (leading to complications, increased length of stay, and potentially worse outcomes) can be as high as \$500,000. Hypothetically, the hospital collects the following data from patients:

1. **Age:** Patient’s age in years.
2. **Gender:** Male (1) or Female (0).
3. **Chest Pain Type:** Typical Angina (1), Atypical Angina (2), Non-anginal Pain (3), Asymptomatic (4).
4. **Smoker:** Yes (1) / No (0).
5. **Resting Blood Pressure:** Systolic blood pressure at rest (mmHg).
6. **ECG Results:** Normal (1), ST-T wave abnormality (2), Left ventricular hypertrophy (3).
7. **Troponin Level:** Peak Troponin I or T level (ng/mL).
8. **Heart Rate:** Resting heart rate (beats per minute).
9. **Exercise-Induced Angina:** Yes (1) / No (0).
10. **ST Depression:** ST depression induced by exercise relative to rest.
11. **High-Risk Label:** High risk (1) or Low risk (0) within 30 days.

Objective: Construct and evaluate two classification models, one using Support Vector Machines (SVM) and another using Neural Networks (NN), to classify patients as either high-risk (likely to require an expensive intervention) or low-risk. For each model, perform the following tasks:

1. Analyze and estimate the potential “error costs” associated with:
 - **False Positives:** Incorrectly classifying a low-risk patient as high-risk, potentially leading to unnecessary interventions.
 - **False Negatives:** Incorrectly classifying a high-risk patient as low-risk, potentially leading to missed necessary interventions.
2. Explain how each model could reduce overall healthcare costs and quantify the potential cost savings compared to the current standard of care, considering the costs of both false positives and false negatives.
3. Employ various techniques (e.g., Kernel selection, parameter tuning, activation function, network topology, and cross-validation) to decrease the total error cost, defined as:

$$\text{Total Error Cost} = \text{Cost of False Negatives} + \text{Cost of False Positives}$$

Report Format

- (1) Title and Team Members
- (2) Background and Objectives
- (3) Experiments Setting
- (4) SVM Model
- (5) NN Model
- (6) Comparison of Results
- (7) Analysis and Discussions
- (8) Attachment of Computer Program/Codes, Inputs and Outputs