MATH 748: Weekly Report

Oct 18-24, 2021

Nayeong Kim (nkim10@sfsu.edu)

9 Multi-Class Classification

9.7 Regularized Discriminant Analysis

(1) Friedman(1989)
A compromise between QDA and LDA. Set $\hat{\Sigma}_k(\alpha) = \alpha\hat{\Sigma}_k + (1-\alpha)\hat{\Sigma}$ for $\alpha \in [0,1]$. α can be chosen by hyper-parameter-adjusting(e.g. cross-validation).

- $\alpha = 0 \Rightarrow$ LDA
- $\alpha = 1 \Rightarrow$ QDA

(2) Regularized LDA: $\hat{\Sigma}(r) = r\hat{\Sigma} + (1-r)\hat{\sigma}^2 I$

(3) Regularized QDA: $\hat{\Sigma}_k(\alpha, Y) = \alpha\hat{\Sigma}_k + (1-\alpha)\gamma\hat{\Sigma} + (1-\alpha)(1-\gamma)\hat{\sigma}^2 I$
We can choose α, γ by cross-validation.

- $\alpha = \gamma = 1 \Rightarrow$ QDA
- $\alpha = 0, \gamma = 1 \Rightarrow$ LDA

9.8 More flexible, more complex decision boundaries

This is mentioned in *Chapter 12*

(1) FDA (Flexible Discriminant Analysis)

(2) PDA (Penalized Discriminant Analysis)

(3) MDA (Mixture Discriminant Analysis)
While LDA sets $X|Y = k \sim \text{MVN}(\mu_k, \Sigma)$, MDA sets $X|Y = k \sim \pi_1\text{MVN}(\mu_{k,1}, \Sigma) + (1-\pi_1)\text{MVN}(\mu_{k,2}, \Sigma)$

10 Remedies for Severe Class Imbalance

Introduction

(1) What is imbalanced data?
Imbalanced data is a classification problem where the classes are not represented equally. (e.g. fraud transaction, online advertising, disease screening, job application, credit card application)

(2) Why is this a problem?
Hard to predict the minority class.

(3) How to resolve this problem.
Resampling

- Under-sampling
  idea
  benefit
  problem: loss of info

- Over-sampling
  idea: replication
  benefit: No loss of information
  problem: Might lead to overfitting

Synthetic Data Generation
Reading: SMOTE

- ROC Curve: Tell you both the type1, type2 error. x axis: type I error rate
y axis: type II error rate
area under the curve = area below the curve ≤ 1 Each corner : extreme cases
Ideal curve:

10.1 ddd

11 Feature Selection

11.1 Motivation

11.2 Feature Selection

It is a process of selection subset of predictors.
(1) Advantages
(2) Applications
(3) 3 types of Feature Selection Methods

- Filter: Give rankings to the data and take top

- Wrapper:

- Embedded:

11.3 Filter

It is independent of classifier (1) Univariate Filter e.g. Y, $X_1 \to P_1$
$Y vs Y_p \to P_p$ smallest p-value ranks the original use correleation

- Pearson's correlation
- ANOVA test
- A lot of tests

(2) Multivariate Filter Evaluate an entire feature subset Slower

11.4 Wrapper

Interacting with the classifier. Wrapper is guided by the performance of the classifier on the subset. Classifier dependent. 3 Highs: cost, chance of overfiting, success 2 major search schemes

- Sequential
  Not guaranteed. Only look at the moment
- Randomized alg.

11.5 Filter vs Wrapper

- Interaction with the classifier
- Speed
- Performance

11.6 Subset Selection

(1) Best subset selection k predictors $\longrightarrow Y = \beta_0 + \beta_1 X + \beta_k$ Select a single best model from among the $p+1$ models. 2^p very learge It is computationally expensive How to decide? Smallest RSS or largest R^2 Credit data: e.g. from ISL book

References

[1] Wikipedia: Hamming Distance