بازشناسی الگو

درس ۱۳

ارزیابی طبقه‌بندی‌کننده

Classifier Evaluation

کاظم فولادی قلعه
دانشکده مهندسی، پردیس فارابی
دانشگاه تهران

http://courses.fouladi.ir/pr

SYSTEM EVALUATION

- The goal is to estimate the error probability of the designed classification system

- Error Counting Technique
  - Let M classes
  - Let N_i data points in class ω_i for testing.

$$\sum_{i=1}^{M} N_i = N \qquad \text{the number of test points.}$$

  - Let P_i the probability error for class ω_i
  - The classifier is assumed to have been designed using another **independent** data set
  - Assuming that the feature vectors in the test data set are independent, the probability of k_i vectors from ω_i being in error is

$$\text{prob}\{k_i \text{ in } \omega_i \text{ wrongly classified}\} = \binom{N_i}{k_i} P_i^{k_i} (1-P_i)^{N_i-k_i}$$

- Since P_i's are not known, estimate P_i by maximizing the above binomial distribution. It turns out that

$$\boxed{\hat{P}_i = \frac{k_i}{N_i}}$$

- Thus, count the errors and divide by the total number of test points in class

- Total probability of error

$$\hat{P} = \sum_{i=1}^{M} P(\omega_i) \frac{k_i}{N_i}$$

- Statistical Properties

  - $E[k_i] = N_i P_i$
  - Thus, $E[\hat{p}] = \sum_{i=1}^{M} P(\omega_i) P_i = P$
  - $\sigma_{k_i}^2 = N_i(1 - P_i)P_i$
  - $\sigma_{\hat{p}}^2 = \sum_{i=1}^{M} P^2(\omega_i) \dfrac{P_i(1 - P_i)}{N_i}$

Thus the estimator is unbiased but only asymptotically consistent. Hence for **small *N*, may not be reliable**

- A theoretically derived estimate of a sufficient number N of the test data set is

$$N \approx \frac{100}{P}$$

Thus, for $P \approx 0.01$, $N \approx 10000$. For $P \approx 0.03$, $N \approx 3000$

❖ Exploiting the finite size of the data set.

➢ **Resubstitution method**:

Use the same data for training and testing. It underestimates the error. The estimate improves for large N and large $\frac{N}{l}$ ratios.

➢ **Holdout Method**: Given N divide it into:

N_1: training points

N_2: test points

$N = N_1 + N_2$

- Problem: Less data both for training and test

➢ Leave-one-out Method

The steps:

- Choose one sample out of the N. **Train** the classifier using the remaining $N - 1$ samples. Test the classifier using the selected sample. Count an error if it is misclassified.
- Repeat the above by excluding a different sample each time.
- Compute the error probability by averaging the counted errors

- **Advantages**:
  - Use all data for testing and training
  - Assures independence between test and training samples
- **Disadvantages**:
  - Complexity in computations high
- Variants of it exclude $k > 1$ points each time, to reduce complexity

➢ Confusion Matrix, Recall and Precision

- *Recall* (R_i). R_i is the percentage of data points with true class label i, which were correctly classified in that class. For example, for a two-class problem, the recall of the first class is calculated as $R_1 = \frac{A(1,1)}{A(1,1)+A(1,2)}$.
- *Precision* (P_i). P_i is the percentage of data points classified as class i, whose true class label is indeed i. Therefore, for the first class in a two-class problem, $P_1 = \frac{A(1,1)}{A(1,1)+A(2,1)}$.
- *Overall Accuracy* (Ac). The overall accuracy, Ac, is the percentage of data that has been correctly classified. Given an M-class problem, Ac is computed from the confusion matrix according to the equation $Ac = \frac{1}{N}\sum_{i=1}^{M} A(i,i)$, where N is the total number of points in the test set.

Take as an example a two-class problem where the test set consists of 130 points from class ω_1 and 150 points from class ω_2. The designed classifier classifies 110 points from ω_1 correctly and 20 points to class ω_2. Also, it classifies 120 points from class ω_2 correctly and 30 points to class ω_1. The confusion matrix for this case is

$$A = \begin{bmatrix} 110 & 20 \\ 30 & 120 \end{bmatrix}$$

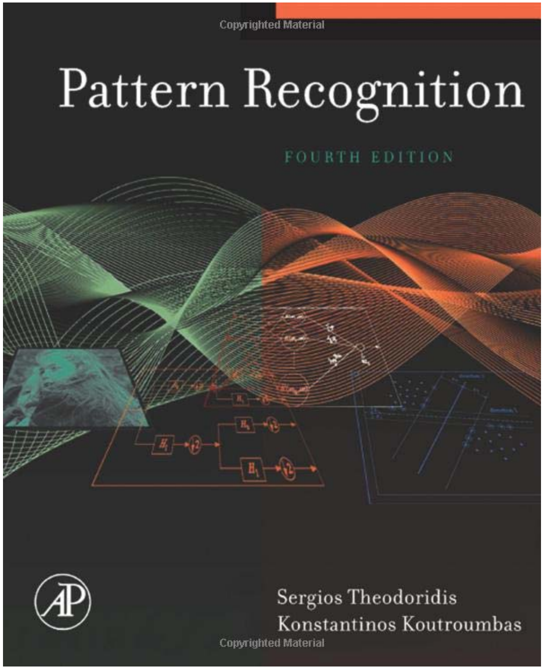
The recall for the first class is $R_1 = \frac{110}{130}$ and the precision $P_1 = \frac{110}{140}$. The respective values for the second class are similarly computed. The accuracy is $Ac = \frac{110+120}{130+150}$.

بازشناسی الگو

ارزیابی طبقه‌بندی‌کننده

منابع

منبع اصلی



S. Theodoridis, K. Koutroumbas,
Pattern Recognition,
Fourth Edition, Academic Press, 2009.

Chapter 10

CHAPTER

Supervised Learning: The Epilogue

10

10.1 INTRODUCTION

This chapter is the last one related to supervised learning, and it is intended to serve three purposes. The first sections focus on the last stage of the design procedure of a classification system. In other words, we assume that an optimal classifier has been designed, based on a selected set of training feature vectors. Our goal now is to evaluate its performance with respect to the probability of classification error associated with the designed system. To this end, methodologies will be developed for the estimation of the classification error probability, using the available, hence finite, set of data. Once the estimated error is considered satisfactory, full evaluation of the system performance is carried out in the real environment for which the system has been designed, such as a hospital for a medical diagnosis system or a factory for an industrial production-oriented system.

It is important to note that the evaluation stage is not cut off from the previous stages of the design procedure. On the contrary, it is an integral part of the procedure. The evaluation of the system's performance will determine whether the designed system complies with the requirements imposed by the specific application and intended use of the system. If this is not the case, the designer may have to reconsider and redesign parts of the system. Furthermore, the misclassification probability can also be used as a performance index, in the feature selection stage, to choose the best features associated with a specific classifier.

The second goal of this chapter is to tie together the various design stages that have been considered separately, so far, in the context of a case study coming from medical ultrasound imaging. Our purpose is to help the reader to get a better feeling, via an example, on how a classification system is built by combining the various design stages. Techniques for feature generation, feature selection, classifier design and system evaluation will be mobilized in order to develop a realistic computer-aided diagnosis medical system to assist a doctor reaching a decision.

In the final sections of the chapter, we will move away from the fully supervised nature of the problem that we have considered so far in the book, and we will allow unlabeled data to enter the scene. As we will see, in certain cases, unlabeled

567