COMP 3605

Rule Based Classifier

Sabina Gooljar

Introduction

In machine learning, particularly in the context of classification, rules are often used to make decisions based on the attributes of the data. Various metrics are used to evaluate the effectiveness of these rules. In this document, we will discuss five such metrics: Accuracy, Coverage, FOIL_Gain, Likelihood Ratio Statistic R, and Laplace Metric.

Evaluation of Rules

a. Accuracy Metric

The accuracy metric measures the proportion of all examples that are correctly classified by a rule.

Rule Accuracy = $\mathrm{n}_{\text{correct}} \frac{}{n_{\text{covers}}}$

A higher accuracy value indicates a better rule.

b. Coverage Metric

The coverage metric measures the proportion of positive examples that are covered by a rule.

The coverage formula is given by:

$$\text{Coverage} = \frac{n_{\text{covers}}}{|D|}$$

A higher coverage value indicates a better rule.

c. FOIL_Gain Metric

The FOIL_Gain metric measures the information gained by adding a new condition to a rule.

$$\text{FOIL\_Gain} = \text{pos'} \times \left(\log_2 \left(\frac{\text{pos'}}{\text{pos'} + \text{neg'}} \right) - \log_2 \left(\frac{\text{pos}}{\text{pos} + \text{neg}} \right) \right)$$

A higher FOIL_Gain indicates a better rule.

d. Likelihood Ratio Statistic R

The likelihood ratio statistic measures the difference between the observed frequencies of the rule's predictions and the frequencies that would be expected if the rule made its predictions randomly.

$$R = 2 \sum_i f_i \log \left(\frac{f_i}{e_i} \right) \tag{1}$$

where m is the number of classes, fi is the observed frequency of class i examples that are covered by the rule, and ei is the expected frequency of a rule that makes random predictions,

A higher value of R indicates a better rule.

Laplace Estimate

In the formula provided for the Laplace estimate:

$$\text{Laplace} = \frac{f + +1}{n + k} \tag{2}$$

- $f+$ is the number of positive examples (or successes) covered by the rule.
- n is the total number of examples covered by the rule (both positive and negative).
- k represents the number of possible class labels (or outcomes).

The Laplace estimate is a form of additive smoothing. The idea is to avoid having zero probabilities when estimating the likelihood of unseen events. By adding a constant (usually 1) to the numerator and adding k to the denominator, we ensure that every class label gets a non-zero probability, even if it hasn't been observed in the data.

In binary classification, where there are only two possible outcomes (e.g., positive and negative), k would be 2. In multi-class classification problems, k would be equal to the number of distinct class labels.

In essence, k in the Laplace formula accounts for the number of possible outcomes or classes in the dataset.

A higher Laplace value indicates a better rule.

Example 1

Consider another training data set D that contains $p = 100$ positive examples and $n = 50$ negative examples. We are given two new candidate rules:

- Rule $r1$: covers $p1 = 60$ positive examples and $n1 = 10$ negative examples.
- Rule $r2$: covers $p2 = 5$ positive examples and $n2 = 1$ negative examples.

$$\text{accuracy} = \frac{n_{\text{correct}}}{n_{\text{covers}}}$$

$$\text{coverage} = \frac{n_{\text{covers}}}{|D|}$$

$$acc(r_1) = \frac{60}{60+10} \qquad acc(r_2) = \frac{5}{5+1}$$

$$\text{coverage}(r_1) = \frac{70}{100+50}$$

$$\text{coverage}(r_2) = \frac{5+1}{100+50}$$

$$\text{FOIL GAIN}(R, R') = pos' \times \left[\log_2\left(\frac{pos'}{pos' + neg'}\right) - \log_2\left(\frac{pos}{pos+neg}\right)\right]$$

$$\text{Foil gain}(r_0, r_1) = 60 \times \left[\log_2\left(\frac{60}{60+10}\right) - \log_2\left(\frac{100}{100+50}\right)\right]$$

$$\text{Likelihood ratio} \quad R = 2\sum_{i=1}^{m} f_i \log_2\left(f_i / e_i\right)$$

$$r_1: \text{expected freq of pos } (e_{pos}) = pos + neg \times \left(\frac{pos}{total}\right)$$

$$= (60+10) \times 100/150$$

$$\text{exp. freq. of neg} = (pos + neg) \times neg/total$$

$$= (60+10) \times 50/150$$

$$R = 2\left[60 \log_2\left(\frac{60}{e_{pos}}\right) + 10 \log_2\left(10/e_{neg}\right)\right]$$

$$\text{Laplace} = \frac{f_+ + 1}{n + K}$$

$$\text{laplace}(r_1) = \frac{60+1}{60+10+2}$$

$$\text{laplace}(r_2) = \frac{5+1}{5+1+2} =$$

Example 2

Consider a training data set D that contains $p = 70$ positive examples and $n = 80$ negative examples. We are given two candidate rules:

- Rule $r3$: covers $p3 = 40$ positive examples and $n3 = 20$ negative examples.
- Rule $r4$: covers $p4 = 4$ positive examples and $n4 = 0$ negative examples.