Naive Bayes Classifier

Paulo S. A. Sousa

2022-04-02

Operating principle

The basic principle for classifying each record is as follows:

1. Find all the other records where the predictor values are the same.
2. Determine what classes the records belong to and which class is most prevalent.
3. Assign that class to the new record.

Cutoff Probability Method

1. Establish a cutoff probability for the class of interest above which we consider that a record belongs to that class.

2. Find all the training records where the predictor values are the same.

3. Determine the probability that those records belong to the class of interest.

4. If that probability is above the cutoff probability, assign the new record to the class of interest.

Conditional Probability

- At the foundation of this method, we find the concept of *conditional probability*:
  - The probability of event A given that event B has occurred, $P(A|B)$.
- We are looking at the probability of the record belonging to class C_i given that its predictor values are $x_1, x_2, \ldots, x_p$, that is:

$$P(C_i|x_1, x_2, \ldots, x_p)$$

- To calculate this probability, we can use the famous Bayes Theorem:

$$P(C_i \mid x_1, \ldots, x_p) = \frac{P(x_1, \ldots, x_p \mid C_i)\, P(C_i)}{P(x_1, \ldots, x_p \mid C_1)\, P(C_1) + \cdots + P(x_1, \ldots, x_p \mid C_m)\, P(C_m)}$$

Practical Difficulty with the Exact Bayes Procedure

- The use of the exact approach requires finding all the records in the sample that are exactly like the new record to be classified in the sense that
  - All the predictor values are all identical.
- When the number of predictors gets larger (even to a modest number like 20), many of the records to be classified will be without exact matches:
  - Even a sizable sample may not contain even a single match for a new record who is:
    - a male Hispanic
    - with high income
    - from the US Midwest
    - who voted in the last election,
    - did not vote in the prior election,
    - has three daughters and one son,
    - and is divorced.

Naive Bayes

The solution is to use naive Bayes, as follows:

1. For class C_1 , estimate the individual conditional probabilities for each predictor $P\left(x_j|C_1\right)$:
    - For X_j this probability is estimated by the proportion of x_j values among the C_1 records in the training set.
2. Multiply these probabilities by each other, then by the proportion of records belonging to class C_1.
3. Repeat Steps 1 and 2 for all the classes.
4. Estimate a probability for class C_i by taking the value calculated in Step 2 for class C_i and dividing it by the sum of such values for all classes.
5. Assign the record to the class with the highest probability for this set of predictor values.

Naive Bayes

In sum, apply the following formula:

$$P_{nb}\left(C_1 \mid x_1, \ldots x_p\right) = \frac{P(C_1)[P(x_1|C_1)P(x_2|C_1)\cdots P(x_p|C_1)]}{P(C_1)[P(x_1|C_1)P(x_2|C_1)\cdots P(x_p|C_1)]+\cdots+P(C_m)[P(x_1|C_m)P(x_2|C_m)\cdots P(x_p|C_m)]}$$

Naive Bayes Assumption of Conditional Independence

- Basically, Naive Bayes assumes, simplistically, that there is conditional independence:

$$P\left(x_1, x_2, \ldots, x_p | C_1\right) \approx P\left(x_1 | C_1\right) \times P\left(x_2 | C_1\right) \times \ldots \times P\left(x_p | C_1\right)$$

- This is not true, as is well-known!

- Despite this probabilistic shortcoming, the algorithm performs quite well in general. ? Why?

Naive Bayes Assumption of Conditional Independence

- The reason is primarily because what is usually needed is not a propensity for each record that is accurate in absolute terms but just a reasonably accurate rank ordering of propensities.
    - Even when the assumption is violated, the rank ordering ofn the records' propensities is typically preserved.

Using the Cutoff Probability Method

- The above procedure is for the basic case, where we seek maximum classification accuracy for all classes.
- When our target class is relatively rare class, we can use the following procedure:

  1. Define a cutoff probability for the target class, above which we consider that a record belongs to that class.
  2. For the class of interest, compute the probability that each individual predictor value in the record to be classified occurs in the training data.
  3. Multiply these probabilities times each other, then times the proportion of records belonging to the class of interest.
  4. Estimate the probability for the class of interest by taking the value calculated in Step 3 for the class of interest and dividing it by the sum of the similar values for all classes.
  5. If this value falls above the cutoff, assign the new record to the class of interest, otherwise not.
  6. Adjust the cutoff value as needed, as an hyperparameter of the model.

Advantages and disadvantages of the Naive Bayes

Advantages

- Simplicity, computational efficiency, good classification performance, and ability to handle categorical variables directly.
- It often outperforms more sophisticated classifiers even when the underlying assumption of independent predictors is far from true.
  - This advantage is especially pronounced when the number of predictors is very large.

Disadvantages

- Naive Bayes classifier requires a very large number of records to obtain good results.
- Where a predictor category is not present in the training data, naive Bayes assumes that a new record with that category of the predictor has zero probability.
  - This can be a problem if this rare predictor value is important.
  - Mitigating solution: Using smoothing, especially Laplace smoothing.

Advantages and disadvantages of the Naive Bayes

- When the goal is to estimate the probability of class membership (propensity), naive Bayes results are very biased.
  - For this reason, the naive Bayes is rarely used in credit scoring (Larsen, 2005).