Naïve Bayes Classifier

The Bayes Theorem gives us information of changes in probability given a particular situation. Bayes theorem act as a basic criteria for many machine learning algorithm and forms the basis of Naïve Bayes classifier. These kind of classifier work in linear time and are very scalable and adoptable. The conditional independence of features may result in wrong classification many times but the ease of use and adaptability of classifiers make these types of classifier as the widespread tool.

We apply the given formulation to our problem in the following manner, using Bayes theorem, we can write the probability of dish belonging to particular cuisine as,

$$P(C_j|d) = P(d|C_j) * \frac{P(C_j)}{P(d)}$$

where $P(d)$ = Probability of occurrence of particular dish. $P(Cj)$ = Probability of occurrence of particular Cuisine. $P(Cj|\ d)$ = Probability of particular dish being in cuisine Cj. (we want to find out this). $P(d|\ Cj)$ = Probability of occurrence of particular dish given a cuisine. (We can find this using training data).

We have assumed here that the occurrence of ingredients is ***not correlated***. This means that the probability of occurring of ingredient is independent of other ingredient present in the dish.

$$P(I_i|I_j) = P(I_i) \quad \forall\ i \in [1, m]\ , j \in [1, m]$$

Second assumption here is probability of occurring of a dish in a cuisine is product of the probabilities of all the ingredients in a dish, i.e. ***dishes are independent***.

$$P(d|C_j) = P(I_1|C_j) * P(I_2|C_j) * P(I_3|C_j) * P(I_4|C_j) \ldots\ldots * P(I_m|C_j)$$

Now with this we can calculate $P(C_j|d)$ and make a classifier. A simple Bayesian classifier will be of the

$$\hat{J} = argmax_{k\epsilon\{1,\ldots.K\}} P(C_k) \prod_{i=1}^{n} P(x_i|C_k)$$

The dish is classified in cuisine which gives the maximum probability.

Now we will see how to do this classification from the data present at hand. From the given training data we can create the document term matrix of the ingredients.

All the ingredients appearing in the Training data

All the Dishes in the training data

Id	Milk	Olive_oil	Salt	Corn	Sugar	Bread	……	Meat
10259	0	0	1	1	0	1		0
25693	0	1	1	1	0	0		1
76420	1	0	0	1	1	0		0
79310	0	0	0	0	0	1		0
…..	1	1	1	0	0	0		0
12831	0	1	1	0	0	0		0

All the matrices in this blog are just representation of original matrices.

Since we know from the training data that particular dish belongs to which cuisine we can sort the data according to given cuisine and take row sum in each particular category of cuisine.

All the ingredients appearing in the Training data

All 20 Cuisine in the training data

Id	Milk	Olive_oil	Salt	Corn	Sugar	Bread	……	Meat
Greek	121	213	1242	231	720	121		98
Indian	723	98	1702	432	931	34		123
Italian	123	213	2753	1231	1231	131		312
Mexcian	312	534	764	67	76	42		98
…..	…	…	….	…	…	…		..
French	234	321	1632	232	324	756		123

Now to convert this matrix to probability matrix we need to make the matrix column stochastic.

All the ingredients appearing in the Training data

All 20 Cuisine in the training data

Id	Milk	Olive_oil	Salt	Corn	Sugar	Bread	……	Meat
Greek	0.121	0.0213	0.1242	0.231	0.0720	0.0121		0.098
Indian	0.072	0.098	0.1702	0.432	0.0931	0.034		0.0123
Italian	0.123	0.0213	0.0275	0.1231	0.1231	0.0131		0.0312
Mexican	0.031	0.0534	0.0764	0.067	0.076	0.042		0.098
…..	…	…	….	…	…	…		..
French	0.234	0.0321	0.01632	0.0232	0.0324	0.0756		0.0123

All the columns sum to one.

Now with this matrix at hand we are in a position to make our Naïve Bayes Classifier. The *i, j*th element of P matrix is $P(I_j|C_i)$.

$$P(d|C_j) = P(I_1|C_j) * P(I_2|C_j)*P(I_3|C_j)*P(I_4|C_j) \ldots..* (I_m|C_j)$$

Using the above equation and Probability matrix calculated we can find the probability of a given cuisine for a particular dish.

$$P(C_j|d) = P(d|C_j) * P(C_j)$$

A given dish is classified to a cuisine which gives maximum probability of belonging to a particular cuisine.

Modifications of Naïve Bayes.

When we are classifying through Naïve Bayes probability matrix, if in training data probability of an ingredient appearing in the cuisine is zero then the whole probability will be zero regardless of other probabilities. Solution can be taking the **Geometric Mean** of the entries in a row which are non-zero. In a situation, if there are only few non-zero entries in a row which will make geometric mean higher than above and in these situations, we can get wrong classifications. To avoid previous situation, we maintained a threshold which is defined as the total number of non-zero ingredients should be greater than particular value and then we have calculated the Geometric Mean (with some minimum number of nonzero terms).

Result obtained from Naïve Bayes and its various variations

Methods	Kaggle score (Accuracy)
Probability multiplication	0.58558
Probability addition	0.44571
Population proportion of cuisines (assuming uniform)	**0.59584**
Modified Naïve Bayes with geometric mean	0.56074
Modified Naïve Bayes allowing minimum number of non-zero terms as 4	0.56114
Modified Naïve Bayes allowing minimum number of non-zero terms as 10	0.56647

The third row in above table correspond to the case where the probability term of $P(Cj)$ is constant for all cuisine i.e. $P(Cj) = 0.05$ for all j in 1 to 20.

It can be seen from the table that the modifications of taking threshold of minimum number of ingredients present in dish to be classified in given cuisine before we take the geometric mean, actually improves the accuracy. The geometric mean and its modifications are less accurate than probability multiplication. The modification of uniform probability gives the best accuracy.