1. Feature process

Firstly, I removed height, weight, gender and age features because of high correlation with EER, P target, F target and C target features. EER and E[g] features were also removed because of high correlation with P, F, C features. Then, instead of using the P[g], F[g], C[g], Salt[g] and four target features directly, I added the following ratio features: Xratio = X[g] / X target, where X means P, F, C and Salt and the added features mean the achieving rate of P, F, C and Salt, respectively.

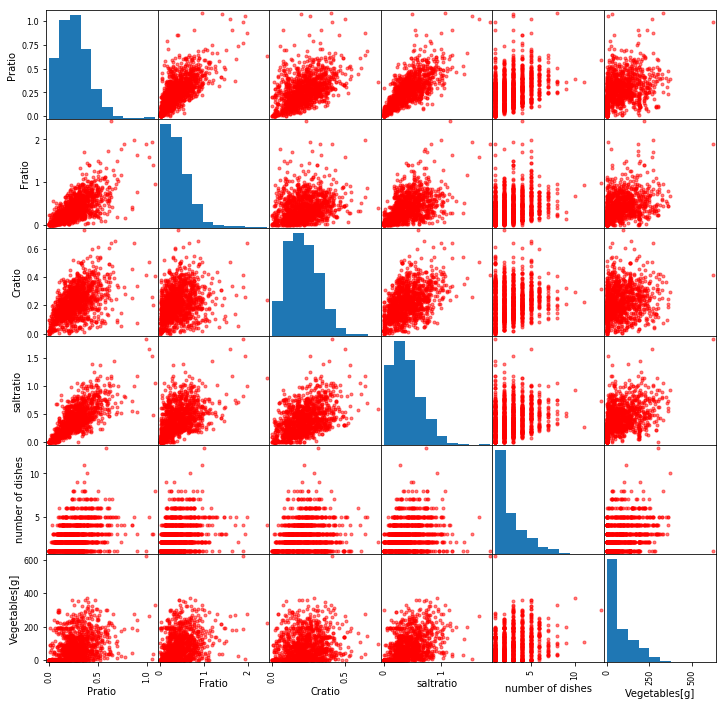
In conclusion, the features I used to train the model are ‘number of dishes’, ‘Vegetables’, ‘Pratio’, ‘Fratio’, ‘Cratio’ and ‘saltratio’.

2. Cross Validation

20% of the train datum were chosen randomly to be as the validation datum for circling 1000 times and the average accuracy of combinations for 1000 times was calculated.

3. Model and accuracy

Firstly, I used the Naive Bayesian Model for breakfast, lunch and dinner types together and the average accuracy was 71%. Then I guessed that the evaluation standards of breakfast, lunch and dinner types are different and I made separate models for them to get better results. The average accuracy of using separate models was about 74%. In order to get better results, I investigated the correlation between each features:



As the figure shows, every two of Pratio, Fratio, Cratio and saltratio features are relevant, which may be explained that the achieving rates of Pratio, Fratio, Cratio and saltratio are all proportional to the amount of meal and increase together. Since the correlation has influence on the Naïve Bayesian Model, I tried to apply the Normal Bayesian Model in which the independent distribution between each feature is not necessary. Normal Bayesian Model thinks that all the features of one class obeys the multiple-variable normal Gaussian distribution,

In the Normal Bayesian Model, the log-likelihood function of conditional probability can be described as followings,

where the , ***x*** and mean the covariance matrices for every class, input vector and mean vector, respectively. The average accuracy of this model was about 75%.

4. Result

At last, the model I used was the Normal Bayesian Model for breakfast, lunch and dinner types, separately. The training datum I used was all the datum except one missing data (1315-1=1314).