In order to verify the good performance of bert, we compare three classifiers with ordinary performance, namely SVM， LogisticRegression and MultinomialNB.

## Logistic Regression

The Logistic Regression is a typical linear classifier and a kind of generalized linear model. It has strong interpretability. The parameters in the Logistic Regression classifier have important meanings. We mainly choose the optimization method **solver** of the logistic regression loss function and the regularization parameter **penalty** to explain.

In the classification task, the use of L1 is to prevent overfitting and unimportant feature coefficients. Our base model has few parameters and is underfitted. So we choose L2 as the **penalty** parameter. The label of this task has only two choices (rumour and nonrumour), so it is a binary logistic regression, which belongs to the scope of OvO processing. Therefore we choose liblinear as the optimization methods while newton-cg, sag, and lbfgs always handle the multinomial loss aka MvM case(Ramos 2005).

LogisticRegression model is prone to underfitting, and the classification accuracy may not be high, poor performance in the face of multivariate or nonlinear decision boundaries.

## Naive Bayesian

Naive Bayesian model assumes that the words are independent of each other, which is often not true in practical applications. And the word statistics in the training set will interfere with the results. The better the training set, the better the results, and the worse the training set, the worse the results.

## SVM

The essence of the support vector machine is to calculate the distance between two observation data. The learning strategy is to maximize the interval, and what it is looking for is the decision boundary that maximizes the sample interval.Since our task this time is linearly separable, there are only two types of labels. We choose LinearSVC (Linear Support Vector Classification)which uses a kernel function "linear".

The parameter C specifies the penalty coefficient of the relaxation factor in the objective function. The larger the C, the greater the penalty for the wrongly classified samples. Base model has low complexity and is under-fitting. At the same time, if C is lower than 1.0, the accuracy will be greatly reduced, so we choose a smaller C of 1.0.

The optional parameters of multi\_class are ovr and crammer\_singer, ovr is mentioned previously one-vs-rest (OvR) or (OVO), and crammer\_singer represents the joint classification strategy, the accuracy will be higher, and the operation process will take a long time. Our training data is only around 2000, so we choose crammer\_singer.

Support vector machines using linear kernel functions are similar to logistic regression, but more robust, and have a certain robustness to test machines, but support vector machines are memory-intensive algorithms, and choosing the correct kernel function requires considerable tricks, not suitable for classification or prediction of large samples.( Hearst 1998)

# Bibliography

I. C. O. Ramos, M. C. Goldbarg, E. G. Goldbarg and A. D. D. Neto, "Logistic regression for parameter tuning on an evolutionary algorithm," 2005 IEEE Congress on Evolutionary Computation, 2005, pp. 1061-1068 Vol. 2, doi: 10.1109/CEC.2005.1554808.

M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt and B. Scholkopf, "Support vector machines," in IEEE Intelligent Systems and their Applications, vol. 13, no. 4, pp. 18-28, July-Aug. 1998, doi: 10.1109/5254.708428.

Text

Description automatically generated

Chart-1