**Noise and incomplete data.**

There are always missing (incomplete) or wrong data in our dataset since these data sets are often recorded and uploaded by sensors or people. For those which contains incomplete data pieces, we need to throw them out or request a complete version. For those which contains wrong data pieces, if we add it to the data analysis in processing the data, the result of our analysis will be much different from the actual situation.

**Algorithm efficiency.**

The speed and accuracy of different algorithms for different problems are different, for example, in sorting problems, the speed of quick sort which average complexity is O(logN) is often faster than bubble sort which average complexity is O(N^2); in small sample analysis prediction, gray prediction will have relatively higher accuracy than regression analysis prediction.

**Incorporation of background knowledge.**

When we perform data mining, we often need to analyze the data in conjunction with the relevant background knowledge of the data as well. With the relevant background knowledge, we can know how to analyze the problem and obtain the data. For example, when determining a person's physical exercise ability, it is often necessary to test the person's blood oxygen level. If we have background knowledge in this area, we can quickly perform relevant modeling based on the blood oxygen level instead of aimlessly searching for relevant factors.

**Complex structured data types.**

When we do data mining, not all data structures are simple, even sometimes there is a certain structure in the string data we get, for example, the URL of a web page contains a lot of data, including protocols, domain names, ports, file names and parameters, etc. These data still have an important role in specific occasions