1. **Why does the probability density function have ‘spikes’? Why is it bad?**

The root cause is limited data. Under the limited data, K and V are also limited. The value of K and V affects the dependence of probability density function on data set. Taking k-nearest neighbor as an example, if K is too large, the dependence of probability density function on data sets will be greater, and it will be more inclined to make judgments from regional considerations rather than specifically considering the situation of a sample itself, so its deviation will be larger and larger. It is equivalent to under fitting. If K is too small, it is easy to be disturbed by noise and wrong samples, which is equivalent to over fitting. The same is true for Parzen window. It can be seen that the value of K or V has a great influence on the probability density function. When the value of K or V is not appropriate, there will be ‘spkies’.

Spkies1 will lead to the decrease of accuracy

1. **When to use the L1 norm and why?**

L1 norm is used when feature selection, compressed sensing or optimal sparse solution is needed. In addition, L1 norm is used to sort of Bag-of-words model.

Because L1 norm can be more sparse than L2 norm, it is easier to get sparse solution when L1 norm is used in cost function.At the same time, L1 norm tends to set the weight of more features to 0, thus completing feature selection or compression sensing.