ML HW1

106598018 萬俊瑋

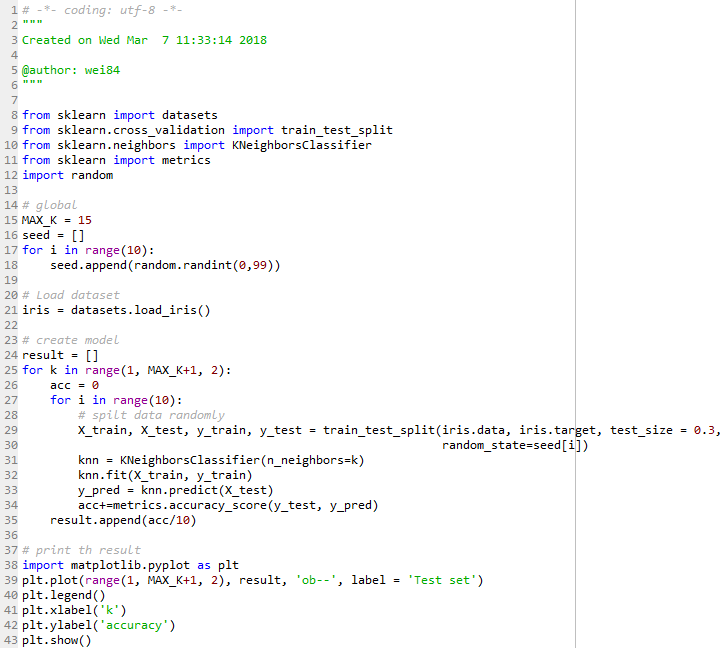
1.

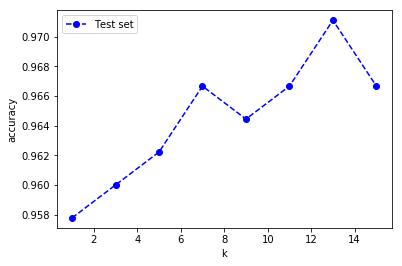
|  |  |
| --- | --- |
| TP 99.9% | FP 0.1% |
| FN 0.1% | TN 99.9% |

For the sicker in group =0.01%

0.01%\*99.9%/(0.01%\*99.9%+99.99%\*0.1%)=0.090834……≒9.1%

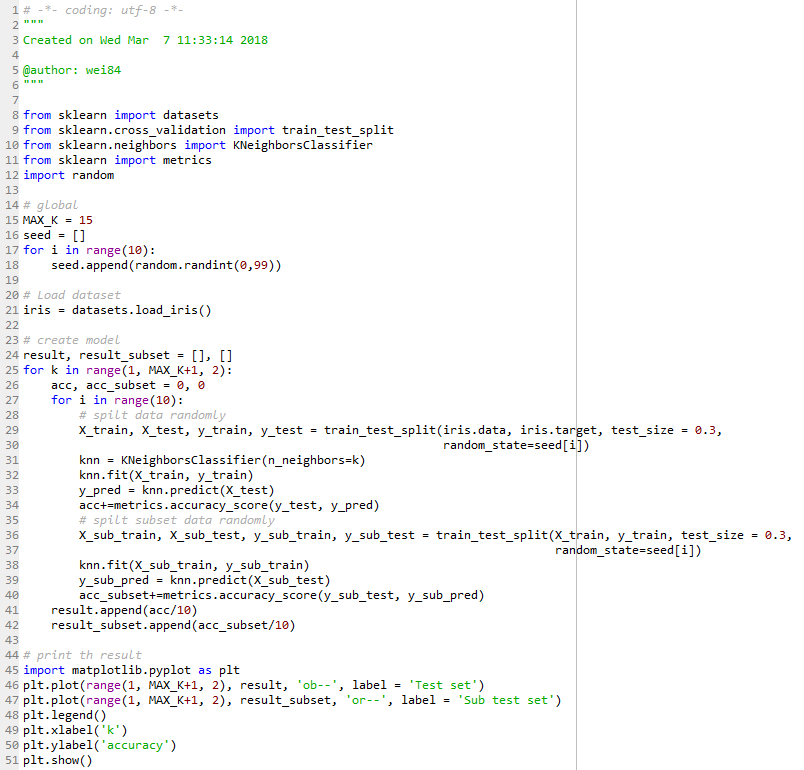
2.

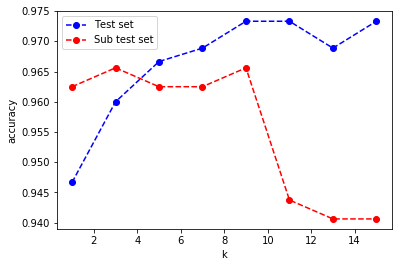




It seems that the best “k” is 13.

3.





I use the 70% subset-data of origin data to spilt into 70% training set and 30% test set again, and run the KNN again to compare the result.

The best “ks” in origin data are 9, 11 and 15, and in subset are 3 and 11.

I run the same code in many times, I found the results are different in every time.

4.

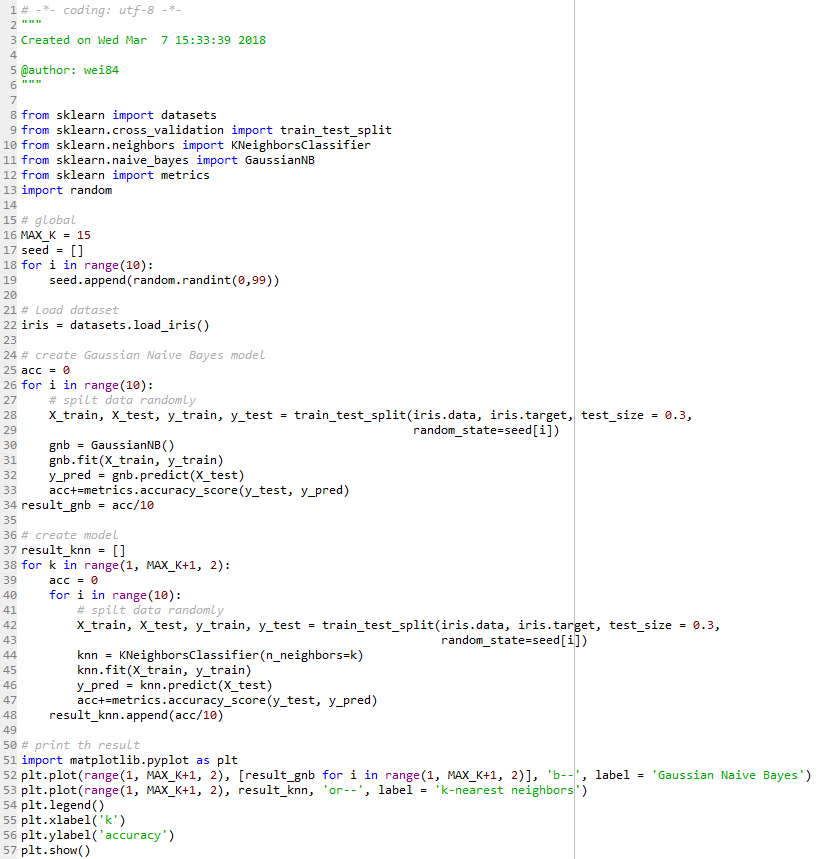
<https://arxiv.org/ftp/arxiv/papers/1302/1302.4964.pdf>

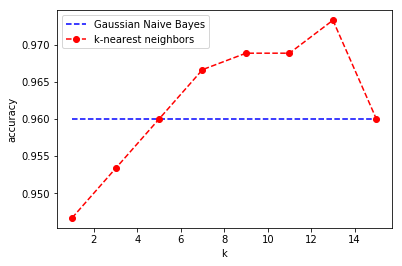
We assume the feather of iris is normal distirbution, so we can do the naïve bayes classify via:

Because we assume iris data is normal distirbution, we can get:

And, we can compute each probility of x in different class.

5.





I found that most of time knn is beter than Gaussian Naive Bayes when k is big enough.