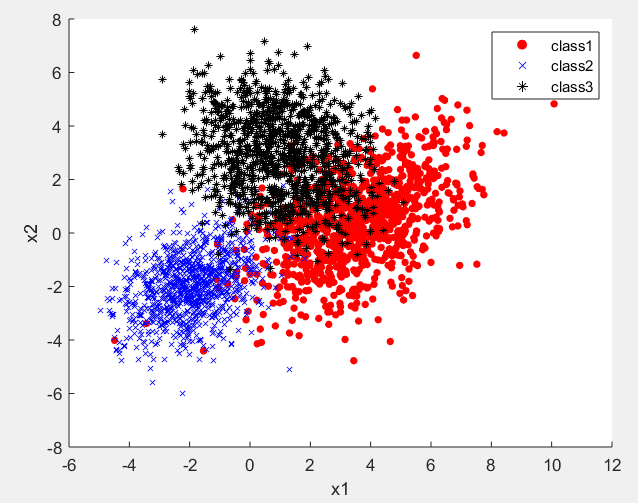
# Problem 1

（a）



（b）

The probability density function formula of binary normal distribution：



[log-likelihood](E:/%E8%BD%AF%E4%BB%B6/%E6%9C%89%E9%81%93%E8%AF%8D%E5%85%B8/Dict/8.9.3.0/resultui/html/index.html" \l "/javascript:;) [function](E:/%E8%BD%AF%E4%BB%B6/%E6%9C%89%E9%81%93%E8%AF%8D%E5%85%B8/Dict/8.9.3.0/resultui/html/index.html" \l "/javascript:;)：



Strives for the partial derivatives：

 and 

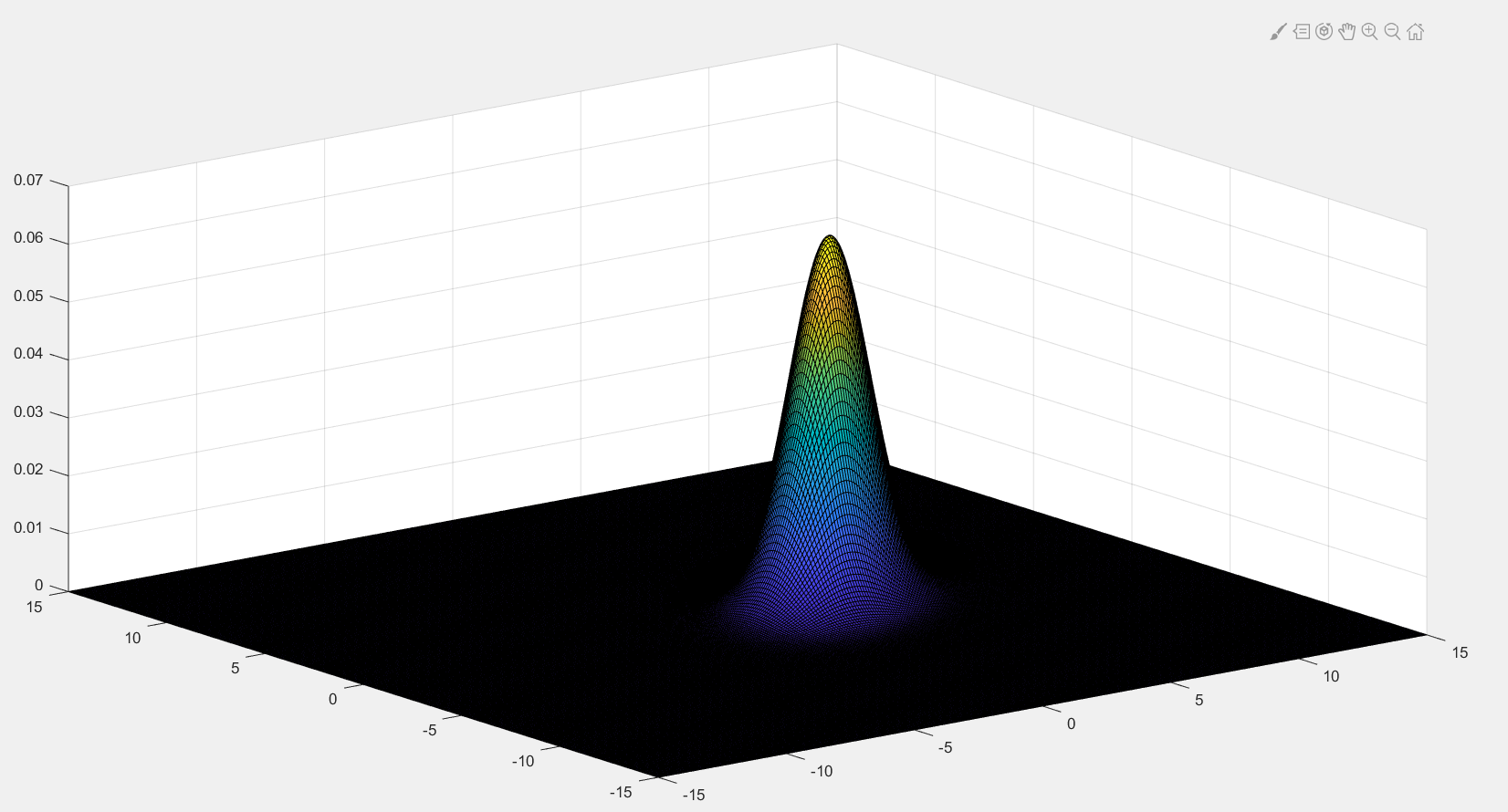
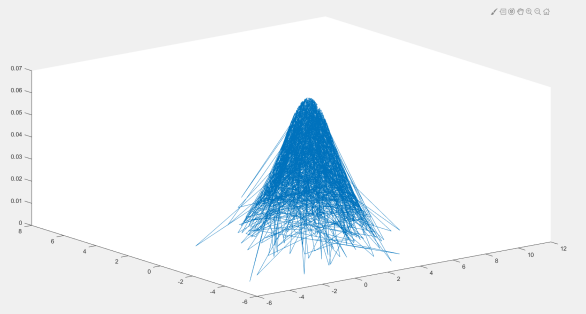
So，we can get:

 and 

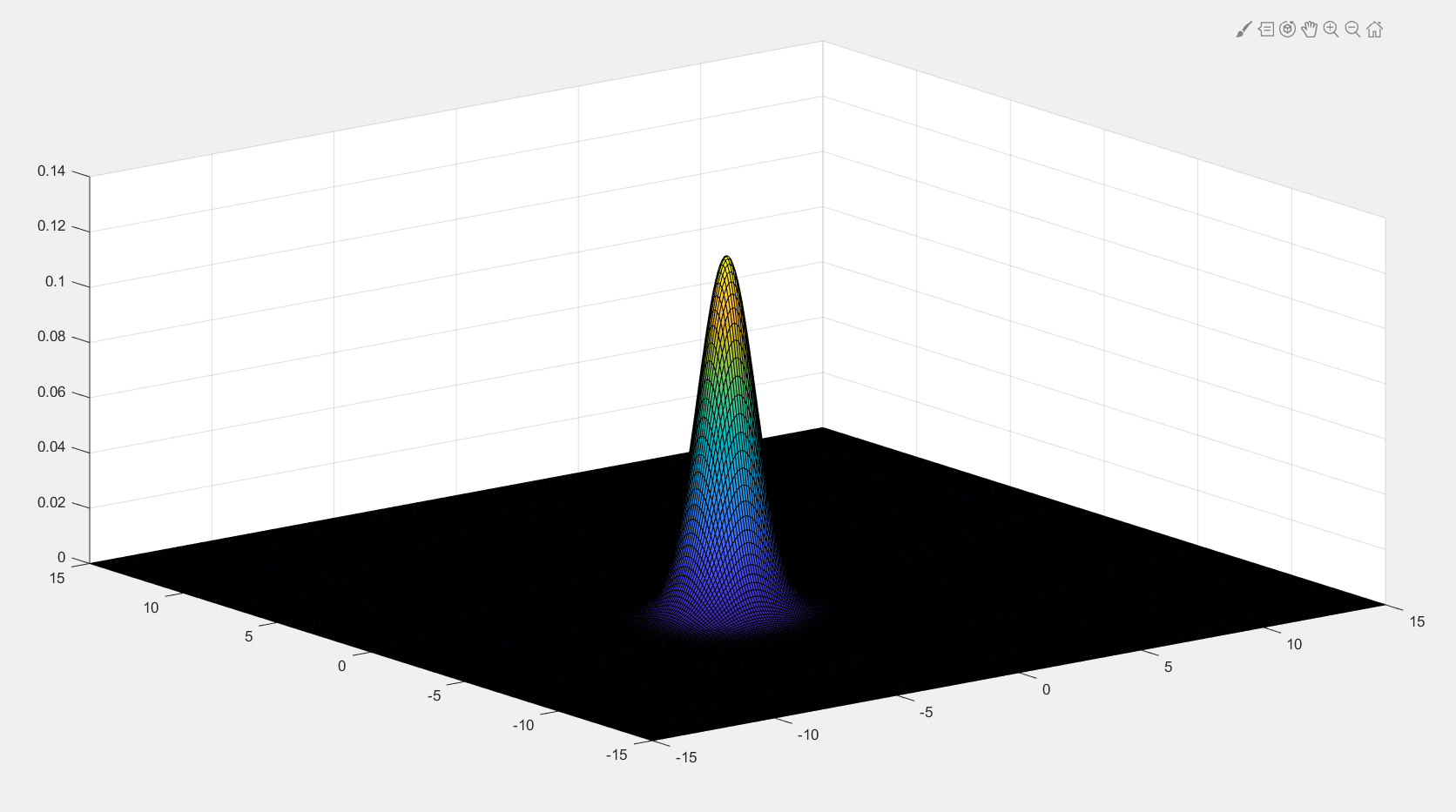
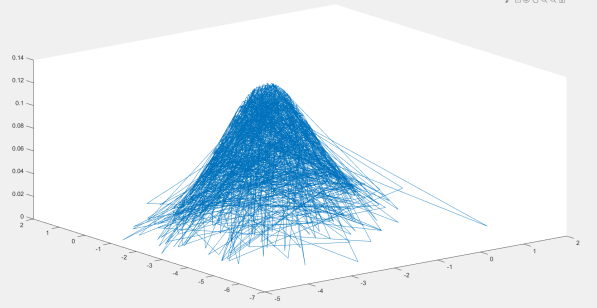
Bring  and  into , we can get the class-conditional densities of each point(Xk).

The difference between the graphs drawn with surf function is not obvious, so I keep the graphs drawn with plot function for the first 1000 points.

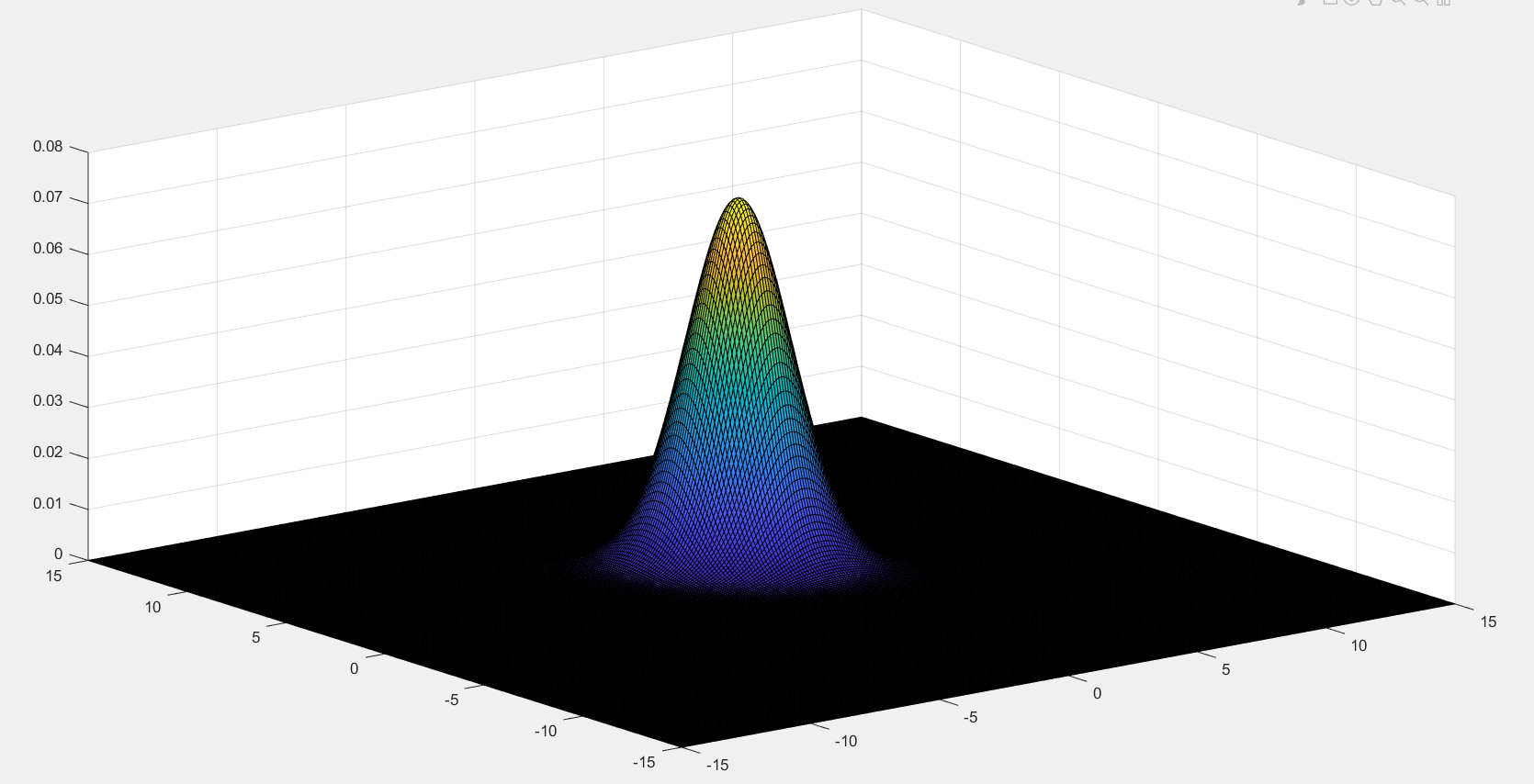
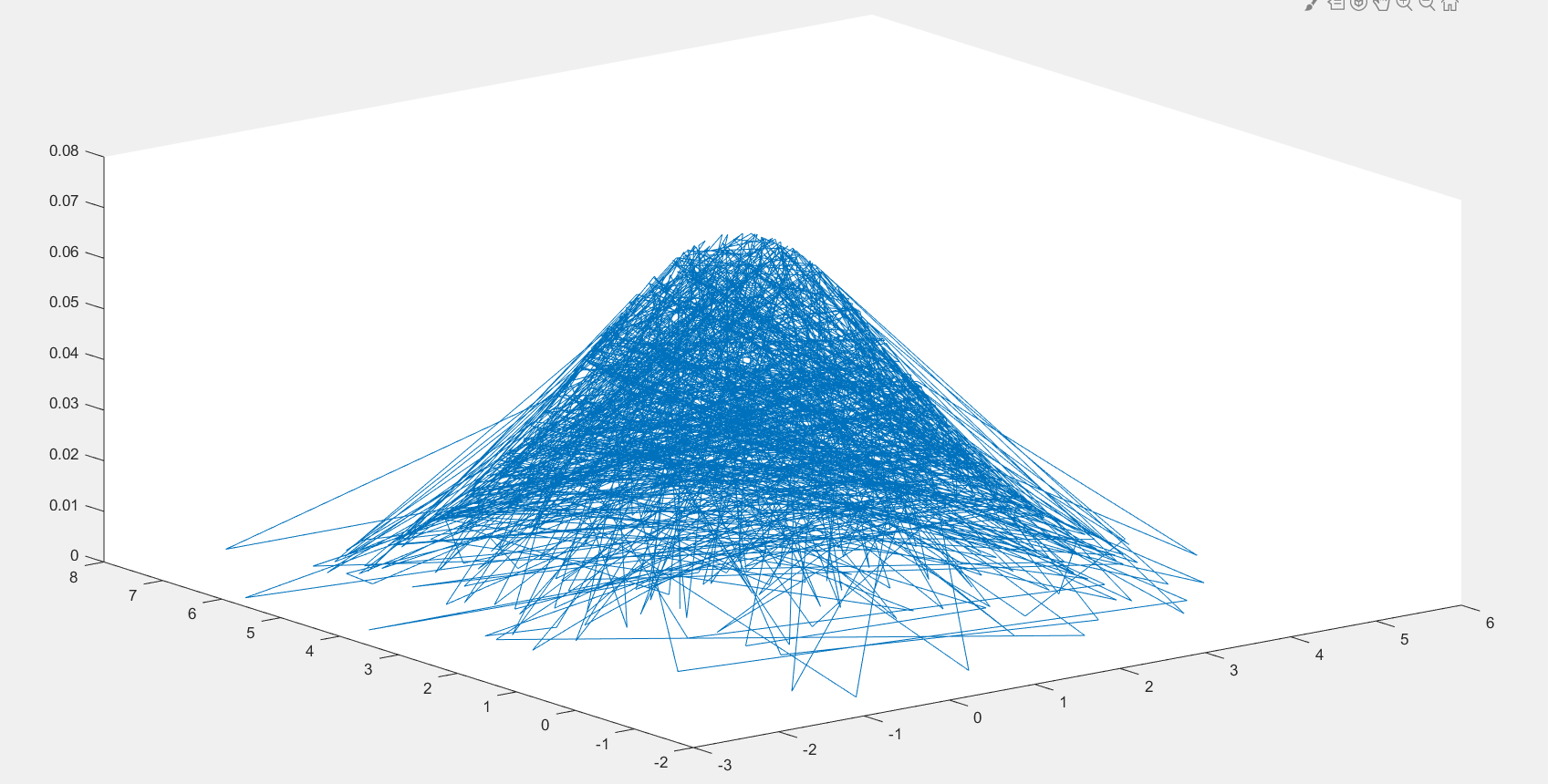
The class-conditional densities of the first data set(plot3 and surf-15:0.1:15)：



The class-conditional densities of the second data set(plot3 and surf-15:0.1:15)：



The class-conditional densities of the third data set(plot3 and surf-15:0.1:15)：



（c）

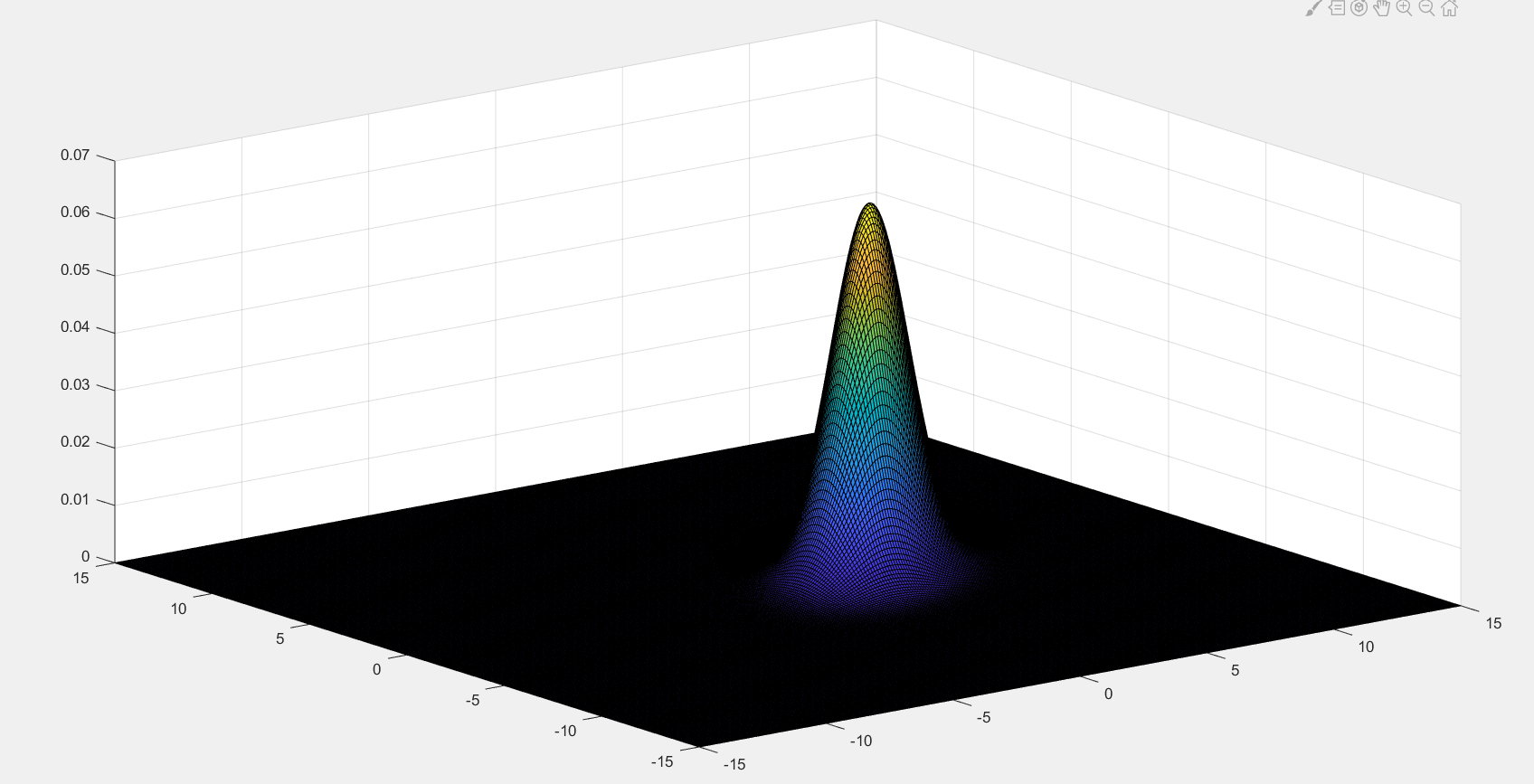
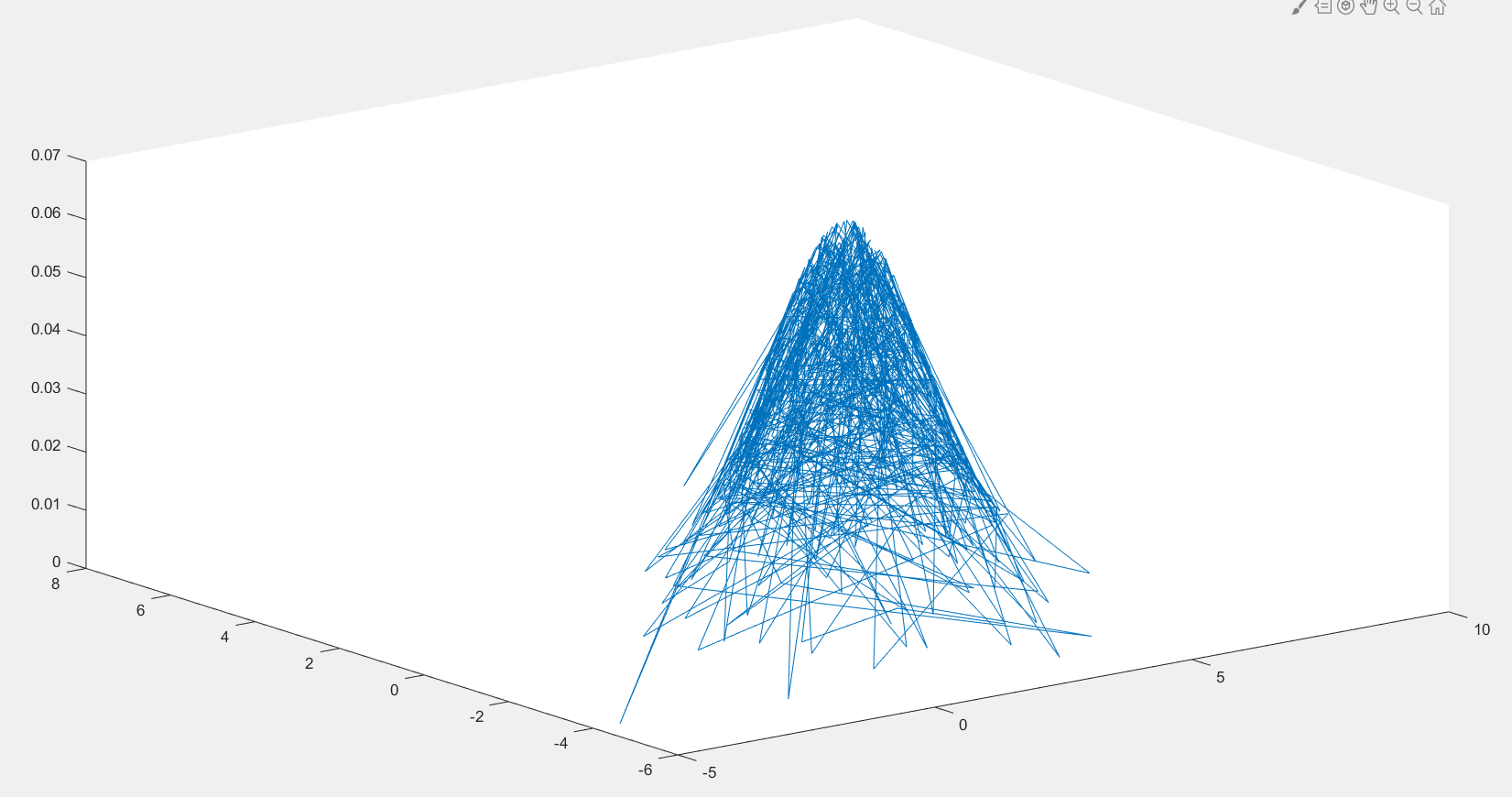
In (b),we get  and 

So, 

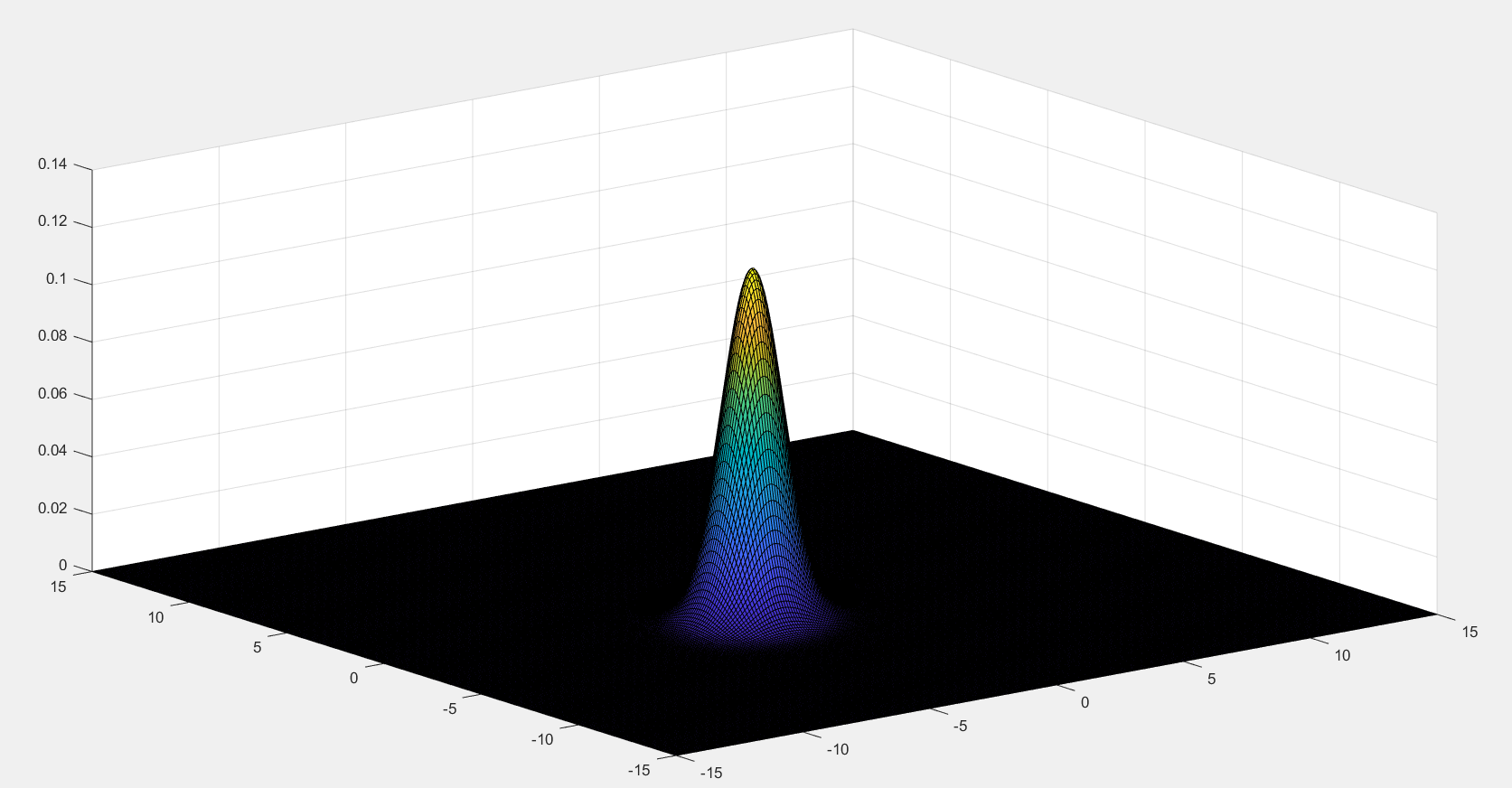
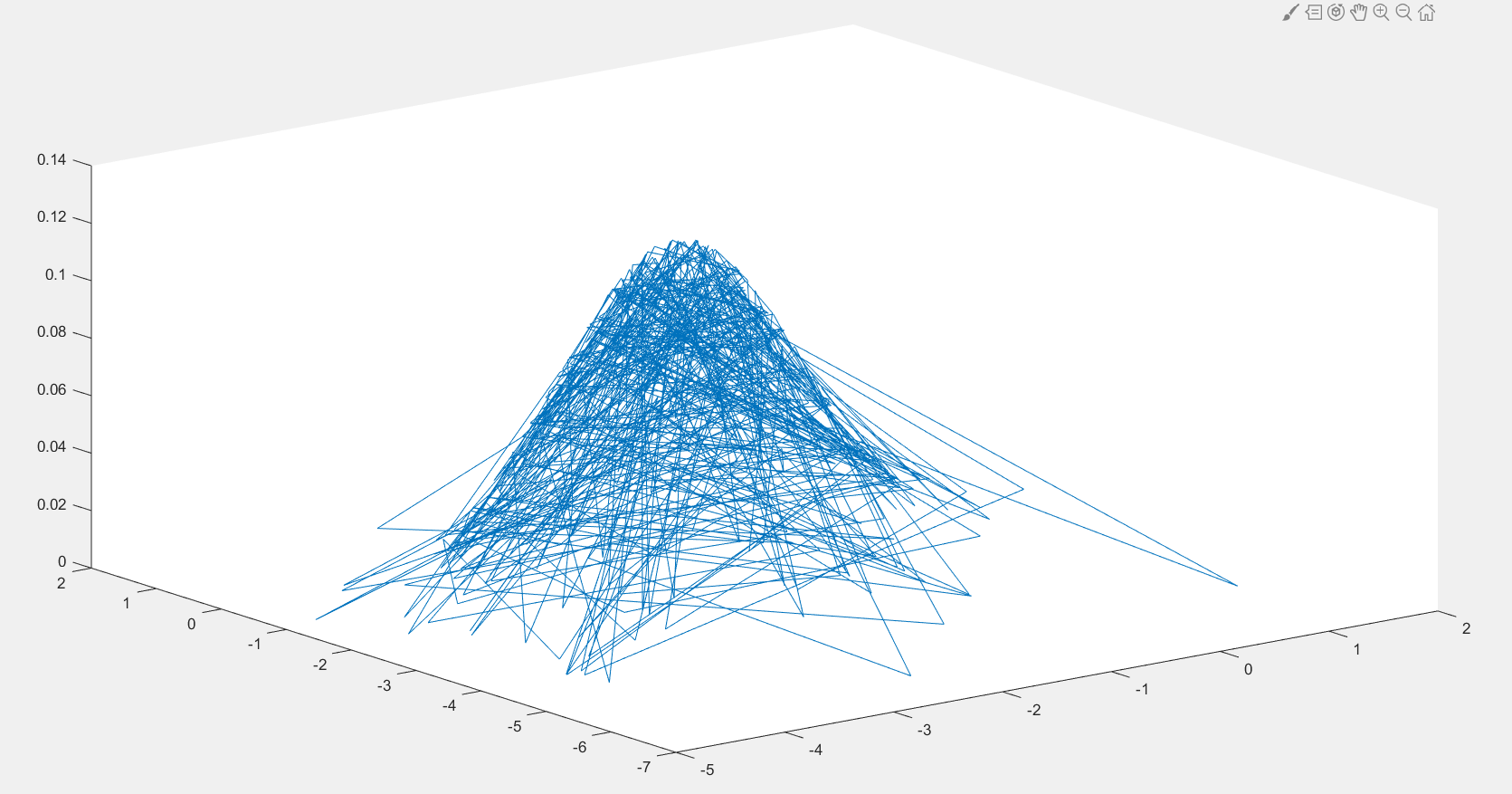
When to use MAP, rate of misclassification in each category：0.1340 0.0250 0.1300

（d）

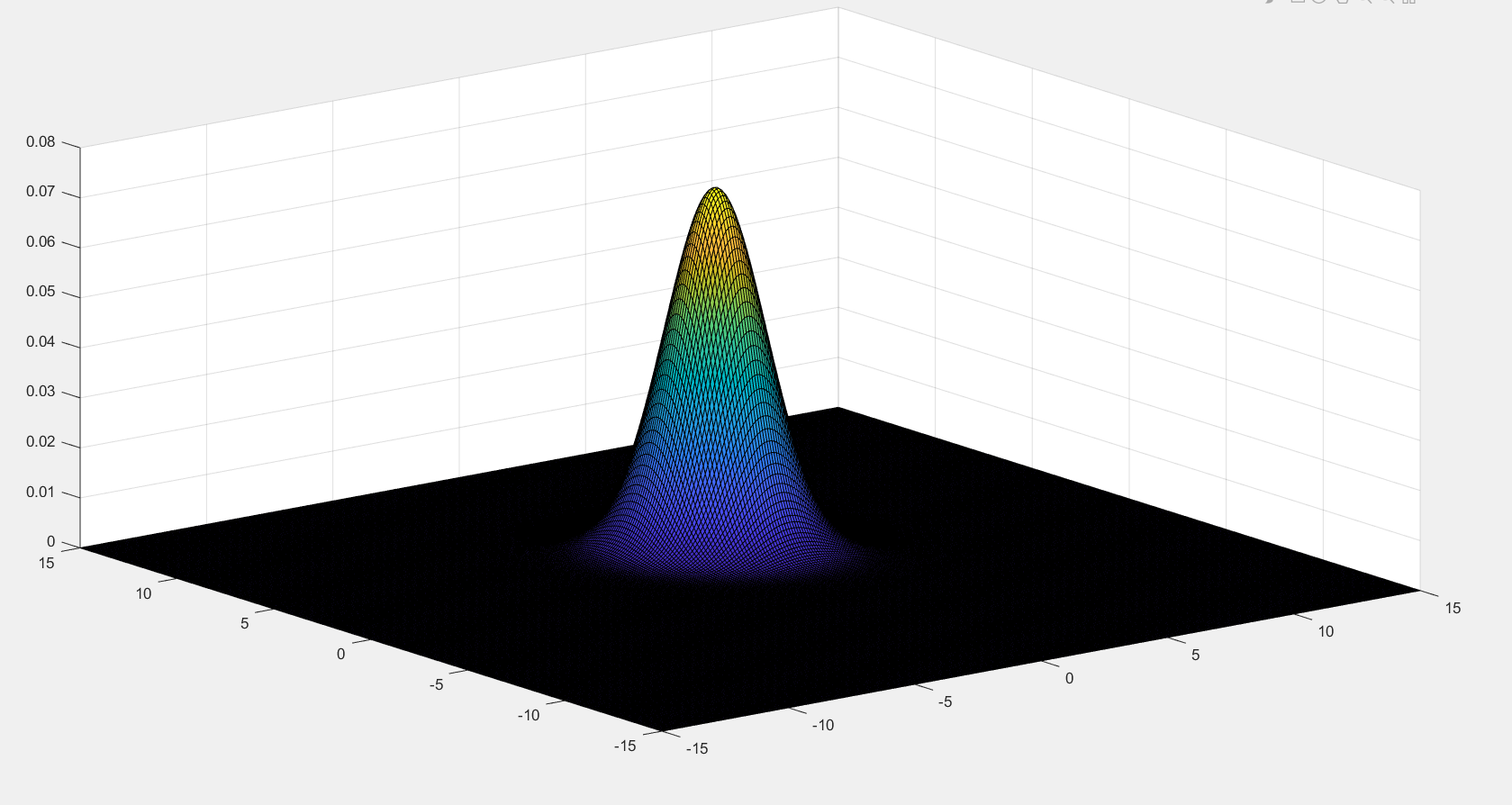
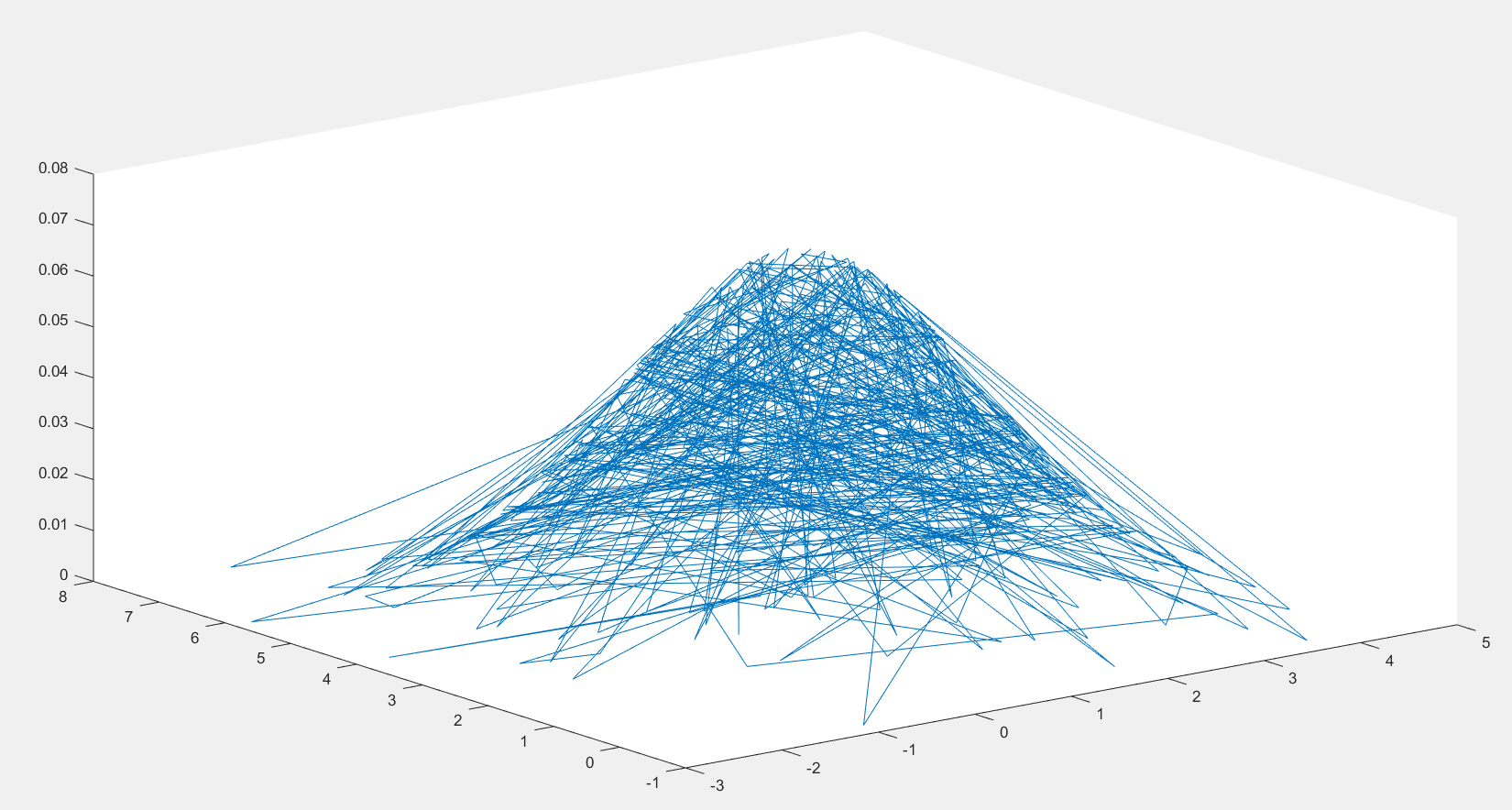
The class-conditional densities of the first data set(plot3 and surf-15:0.1:15)：



The class-conditional densities of the second data set(plot3 and surf-15:0.1:15)：



The class-conditional densities of the third data set(plot3 and surf-15:0.1:15)：



Rate of misclassification in each category：0.3860 0.0740 0.4100

（e）

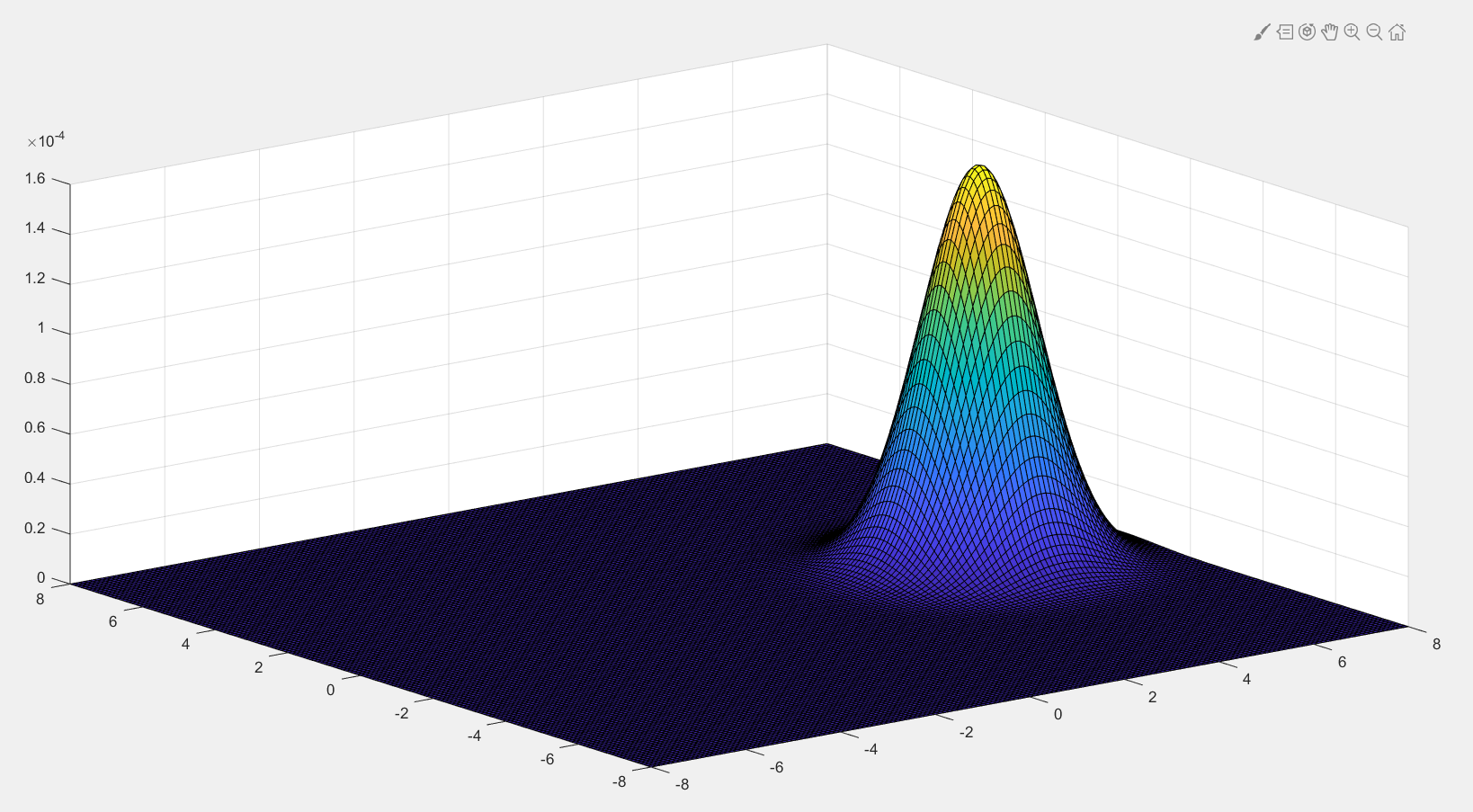
The solution of MLE is only an estimate, only when the number of training samples tends to infinity, it will be close to the real value.This can be seen from the graphs drawn with plot3 function.

# Problem 2

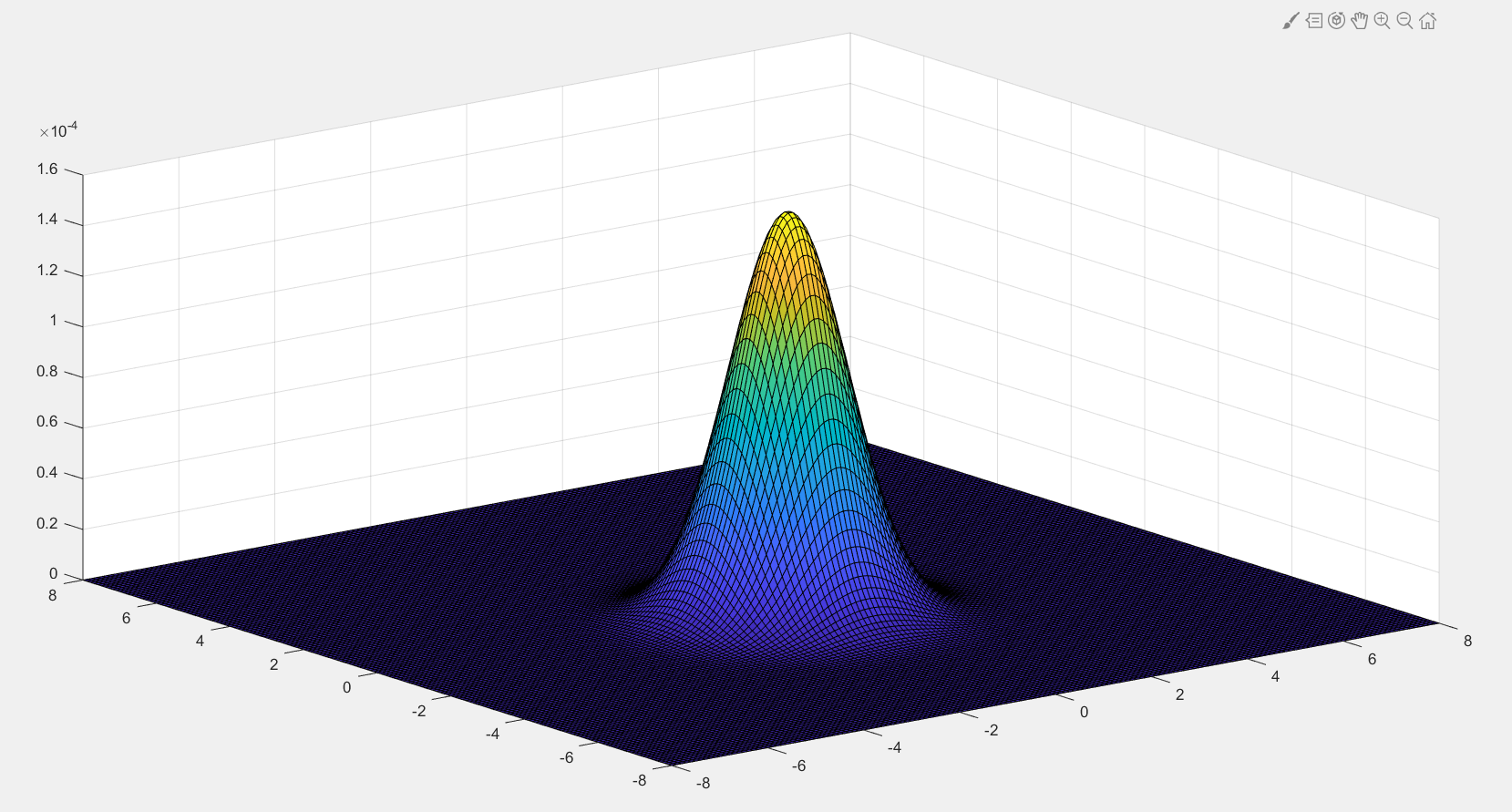
（a）

Just calculate according to the formula given in (a), and we will get the class-conditional densities of each point.

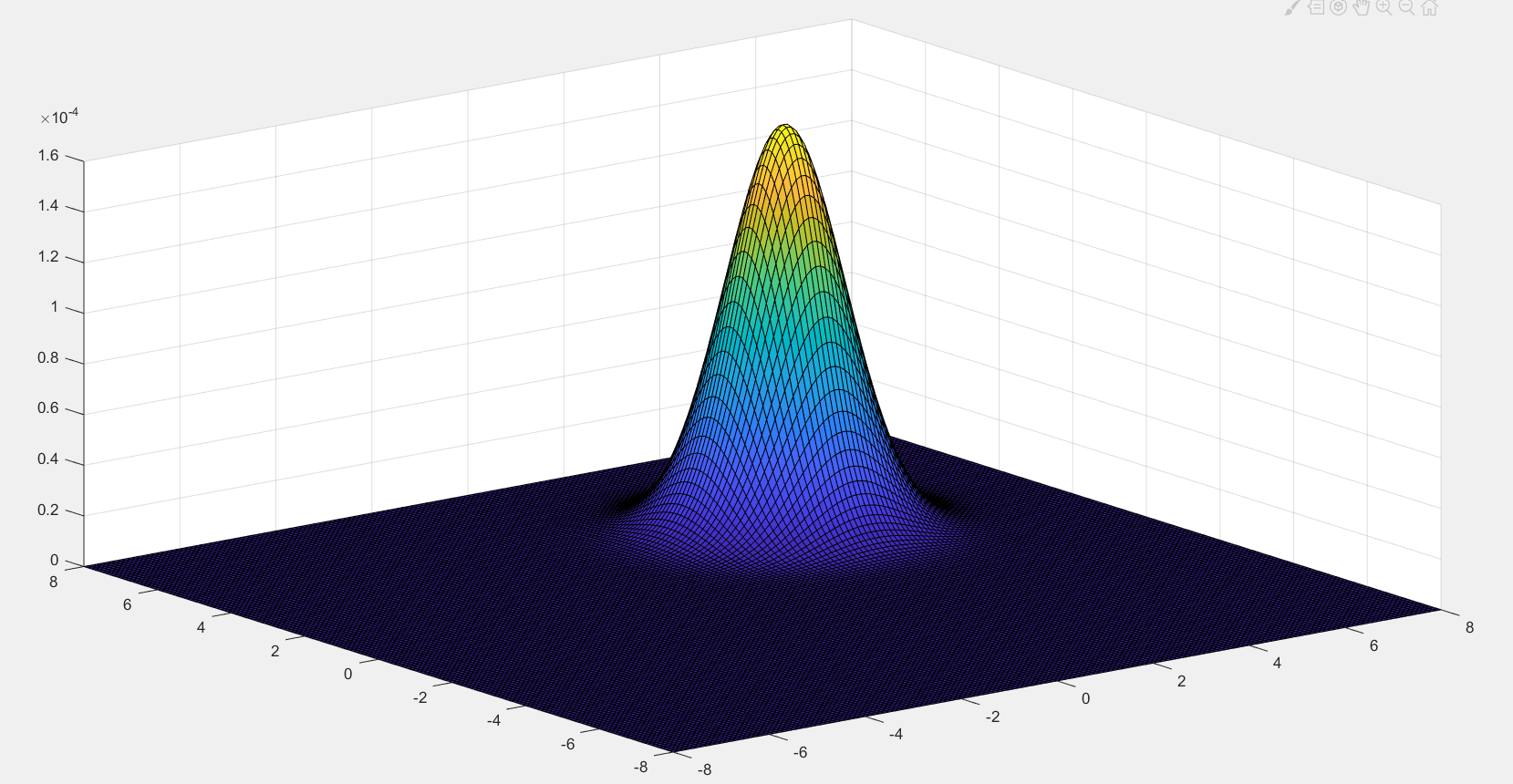
The class-conditional densities of the first data set：(-8:0.1:8)



The class-conditional densities of the second data set：(-8:0.1:8)



The class-conditional densities of the third data set：(-8:0.1:8)



（b）

Whether it's Parzen window or KNN，we can get an estimate of 

And the posterior probability is estimated to be 

When to use MAP, rate of misclassification in each category：0.488000 0.001000 0.197000

（c）

Rate of misclassification in each category：0.193000 0.049000 0.169000

（d）

Rate of misclassification in each category：0.142000 0.025000 0.150000

（e）

The selection of K value in KNN has a great influence on the result of k-nearest neighbor algorithm.

If a smaller K value is selected（like NN in (c)）, it is equivalent to forecasting with a smaller training instance in the field. The "learning" approximation error will be reduced. Only a training instance that is close to or similar to the input instance will play a role in the prediction result. At the same time, the problem is that the "learning" estimation error will increase. In other words, the reduction of K value means that the overall model becomes complex, It is easy to be over fitted;

If we choose a larger K value（like KNN in (d)）, it is equivalent to using the training examples in a larger field to predict. Its advantage is that it can reduce the estimation error of learning, but its disadvantage is that the approximate error of learning will increase. At this time, the training instance which is far away from the input instance will also act on the predictor, making the prediction error, and the increase of K value means that the overall model becomes simple.