ECE 283: Homework 1

Yang Zhao 7497563

yang_zhao@umail.ucsb.edu

Homework group: Yulin Ou, Fanqi Meng

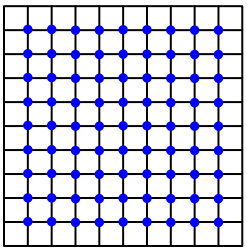
1. Classification by using MAP (Question #1, #2, #3)

MAP Rule for Q#2 :

$$\hat{z} = p(C_1|x) - p(C_0|x) = p(x|C_1)P(C_1) - p(x|C_0)P(C_0) = p(x|C_1) - p(x|C_0)$$
$$if\ \hat{z} \geq 0, Class1; else\ Class0$$

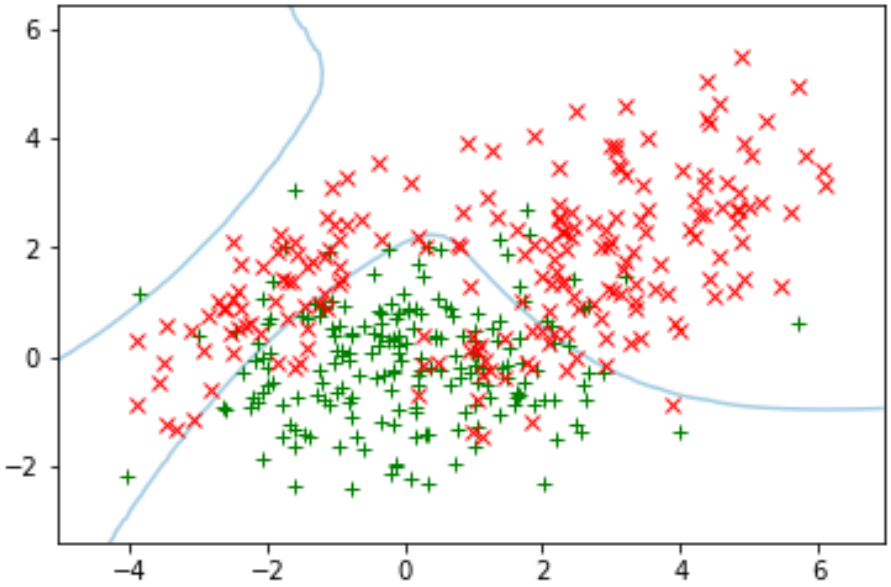
Steps to compute the decision boundary :

Step1: Generate a series of (x_0, x_1) which are eventually scattered on the $x_0x_1 - plane$, like the picture below



Step2: Calculate the corresponding $\hat{z}$ for each (x_0, x_1) by using the MAP Rule show before

Step3: Use plt.contour function to plot the boundary with contour=0. The detailed method is that plotting (x_0, x_1) when the corresponding $\hat{z} = 0$



Estimate the error probability :

Use the MAP Rule show before to calculate the estimated classification. Calculate the number of data whose estimated classification is different from the labeled classification. Divide the number by the total number of data to estimate the error probability.

error_ratio = 15%

2. **Using kernelized logistic regression (Question #4, #5, #6)**

Choose total sample numbers N :

hyperparameter l=1; the number of Newton's method iterations Cycles=10

Sample numbers N	20	50	100	200	400	800
error_ratio (%)	17.8~32.2	16.7~25.5	16.5~22	31.5~62.2	37.7~64.5	29.7~66

The number of training data is set to 100, which can provide lower and more stable error ratio. A proper number of training data means proper dimensions of features. If the dimensions are too high, overfitting occurs easily. If the dimensions are too low, underfitting occurs easily.

Choose hyperparameter l for calculating Gaussian kernel :

The number of total sample numbers N=100; the number of Newton's method iterations Cycles=10

hyperparameter l	0.1	0.5	1	1.5	2	3
error_ratio (%)	17~26.5	17.5~25.5	16.5~22	18~25	20.2~50	29~50

Hyperparameter is set to 1. If hyperparameter l is bigger than 1, underfitting occurs, and the corresponding error ratio is close to 50% when l is bigger than 2. If hyperparameter l is smaller than 1, overfitting becomes more serious.

Choose the number of Newton's method iterations (Cycles) :

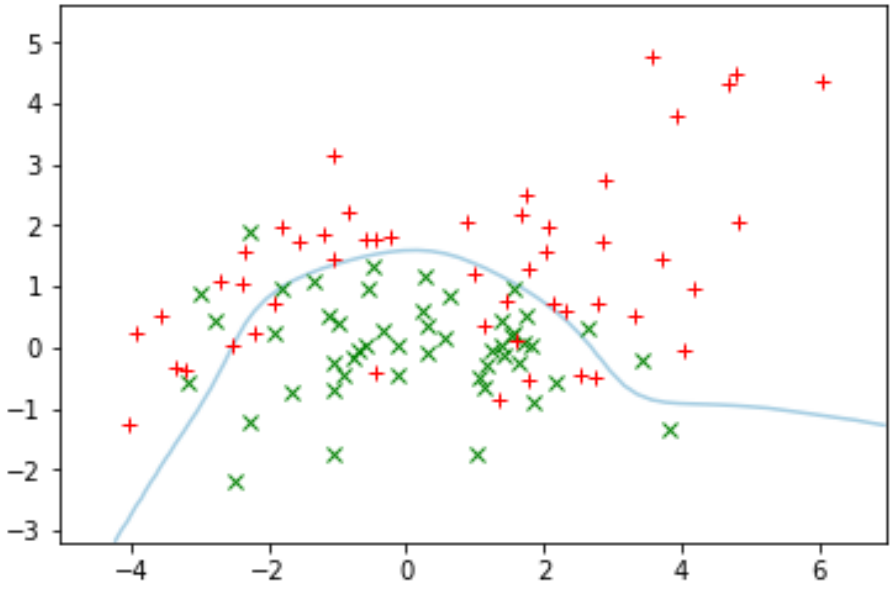
The number of total sample numbers N=100; hyperparameter l=1

Iteration cycles	1	5	10	20
error_ratio (%)	17~22.5	15.5~21.5	16.5~22	17.2~21.2

It seems that the number of iteration cycles does not affect performance.

The decision boundaries :

The steps for kernelized logistic regression is almost the same as for MAP. The only difference is the way to calculate $\hat{z}$ for each (x_0, x_1). Here the non-kernelized logistic regression is used for calculate, which is $\hat{z} = a^T K(X, x_i)$.



Estimate the error probability :

The error probabilities listed in previous tables are computed by using the samples as the MAP method. The error probabilities for kernelized logistic regression are around 15.5%~22% (N=100, l=1).

3. **Using non-kernelized logistic regression (Question #7)**

Choose total sample numbers N :

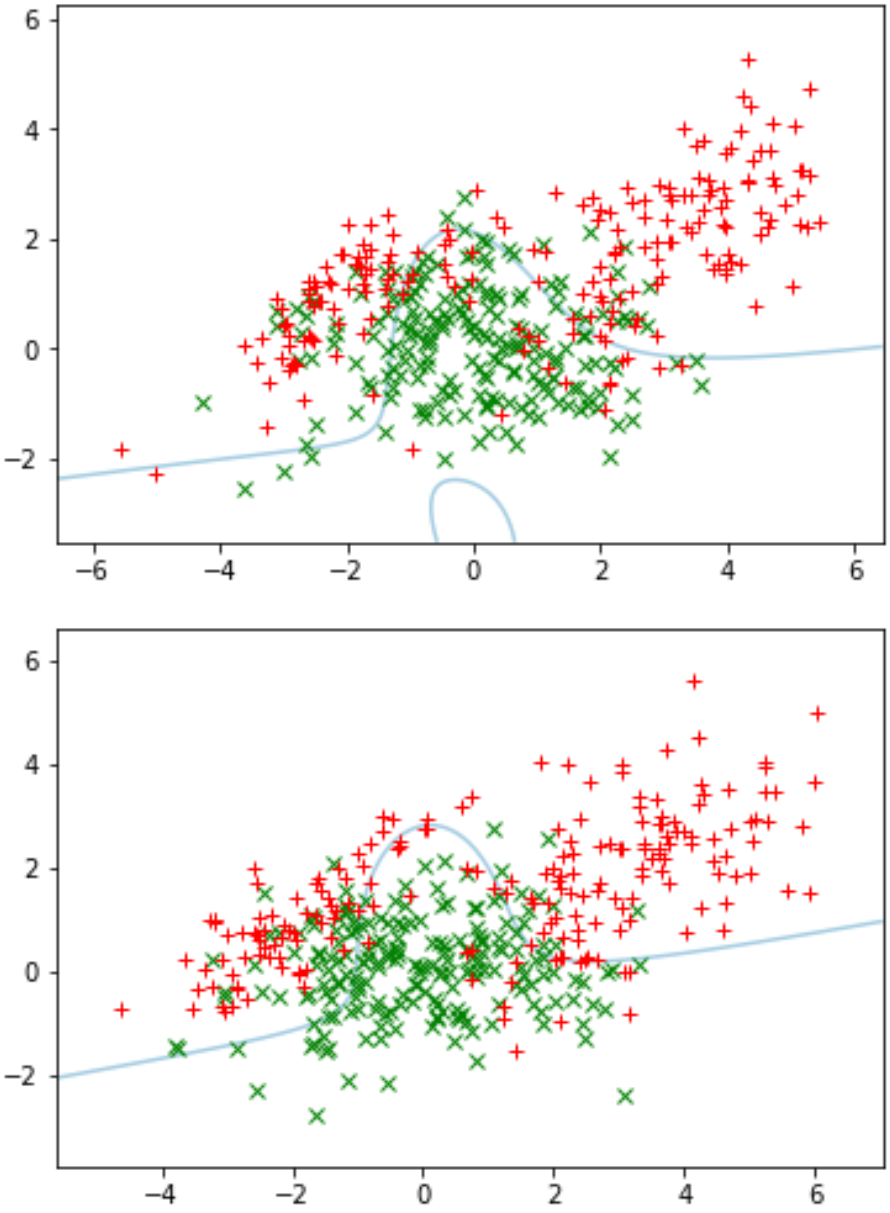
the number of Newton's method iterations Cycles=10; hyperparameter l2=1

Sample numbers N	20	50	100	200	400	800
error_ratio (%)	21.7~36	19~26	18~25.5	16.5~21.5	16.5~18.7	17~18

If the number of sample data is bigger, the error ratio is lower and more stable.

Choose hyperparameter l2 for calculating Gaussian kernel :

the number of Newton's method iterations Cycles=10; the number of sample data is 400.



The upper picture is the case where hyperparameter l2 is 0 with an error probability of 16.5%. The lower picture is the case where hyperparameter l2 is 1 with an error probability of 16.75%. Hyperparameter l2 regularization can help to solve overfitting and reduce computation complexity, but the error probability may become a little worse.

The decision boundaries :

The steps for kernelized logistic regression is almost the same as for MAP. The only difference

is the way to calculate $\hat{z}$ for each (x_0, x_1). Here the kernelized logistic regression is used for calculate, which is $\hat{z} = w^T \Phi(x)$.

Estimate the error probability :

The error probabilities listed in previous tables are computed by using the samples as the MAP method. The error probabilities for non-kernelized logistic regression are around 16.5%~18.7% (N=400, l2=1).

Comparison of kernelized and non-kernelized logistic regression :

	kernelized	**Non-kernelized**
Decision boundary	Decision boundary is easier to become overfitting. Hyperparameter l2 regularization is necessary.	Overfitting is less likely to occur if the dimension of features is selected properly.
Convergence	Because Newton method is used, the number of iteration cycles does not affect performance visibly.	
Misclassification probability	Kernelized logistic regression has better minimum error probability. However, the error probability varies.	Non-kernelized logistic regression has a little worse minimum error probability. However, the error probability is more stable.
Advantages	Need much less number of sample data	The performance is more stable.