Homework 3

Wenjun Zhang

A53218995

For every dataset and each strategy, we are asked to compute the covariance of the class-conditional, and the posterior mean μ_i, and covariance Σ_i of

$$P_{\mu|T}(\mu|D_i) = G(\mu, \mu_i, \Sigma_i)$$

And plot the curves of classification error as a function of α corresponding to the predictive distribution $P_{x|T}(x|D_i)$. Then, we are also required to plot the solutions for MAP and ML.

The posterior parameters μ_i and σ_i can be expressed as:

$$\mu_i = \frac{n\sigma_0^2}{\sigma^2 + n\sigma_0^2}\mu_{ML} + \frac{\sigma^2}{\sigma^2 + n\sigma_0^2}\mu_0$$

$$\frac{1}{\sigma_i^2} = \frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}$$

where σ^2 is the variance of the dataset and μ_0, σ_0 are the mean and variance provided in the strategy. Hence, $x|T$ is Gaussian with $P_{x|T}(x|D_i) = G(x, \mu_i, \sigma^2 + \sigma_i^2)$.

The plots of curves for the four datasets in strategy under are shown as below:

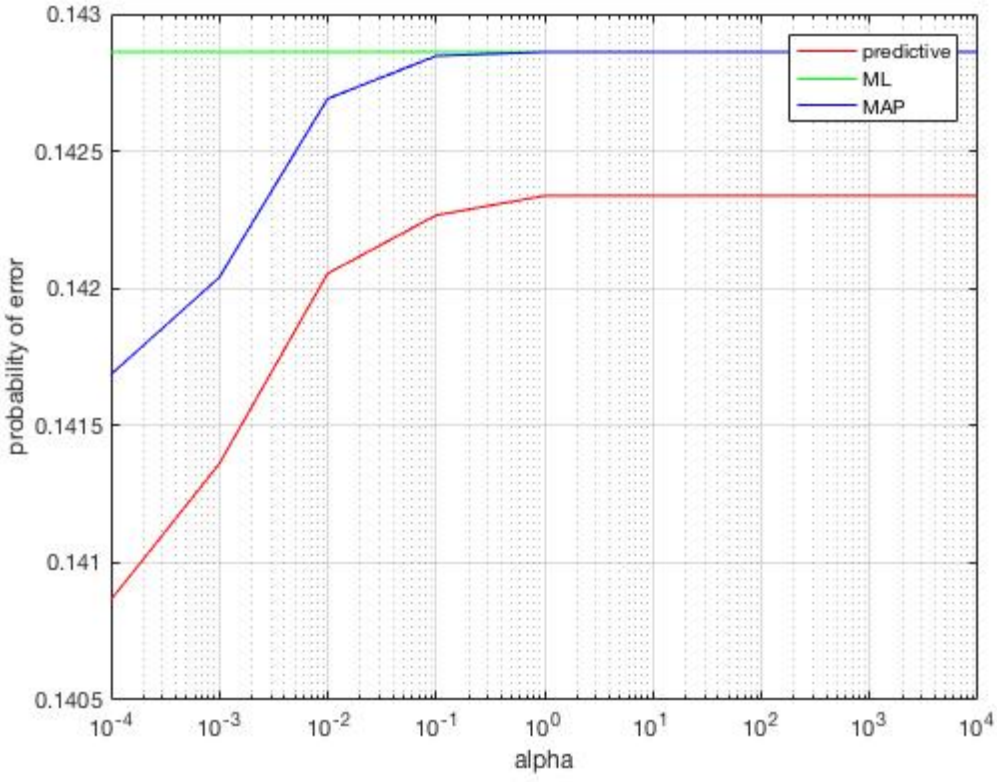

Figure.1 D1 & Strategy 1 Classification Error vs. α

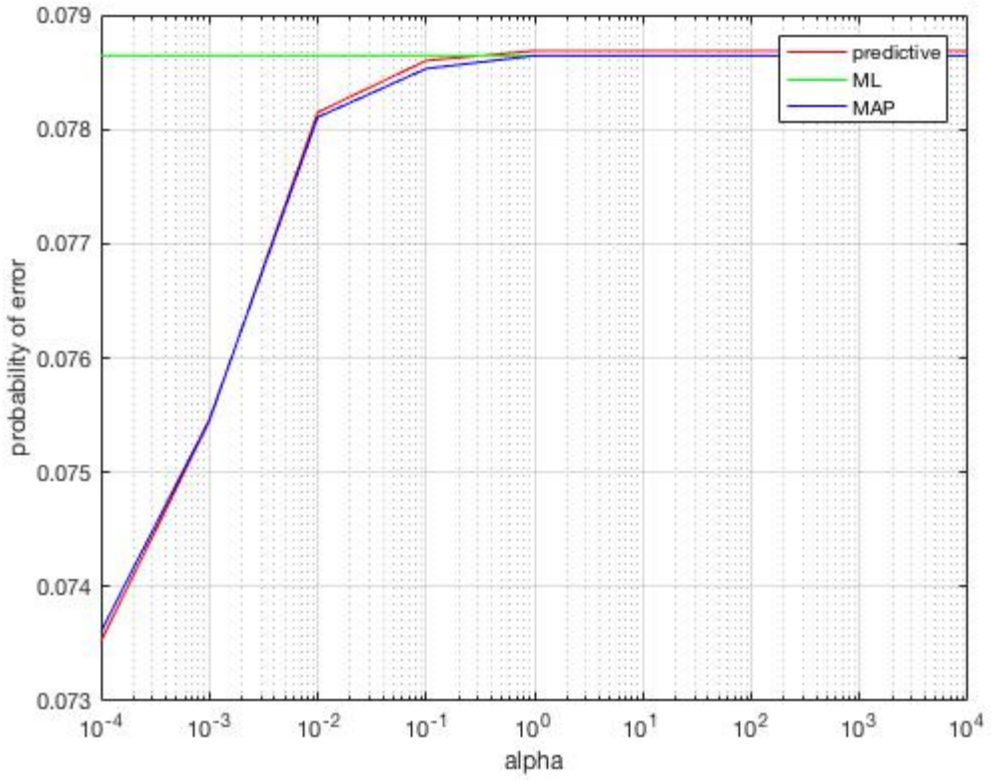

Figure.2 D2 & Strategy 1 Classification Error vs. α

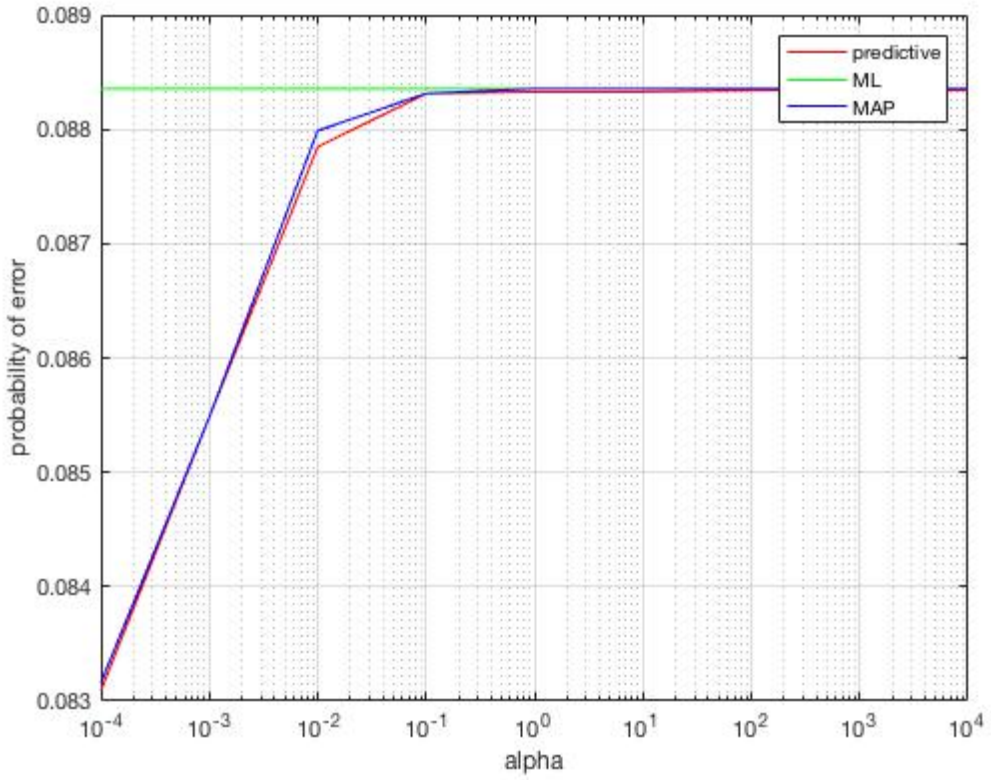

Figure.3 D3 & Strategy 1 Classification Error vs. α

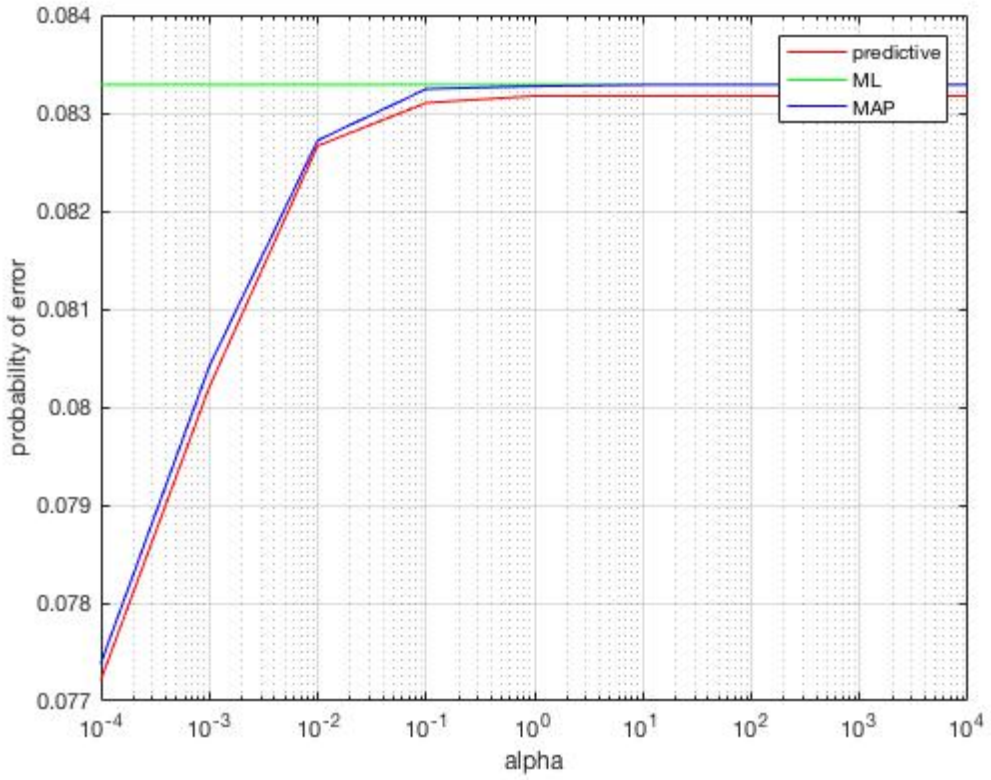

Figure.4 D4 & Strategy 1 Classification Error vs. α

We can see that under strategy 1, for the four plots the classification error of predictive solution increases with the increasing α, and it finally converge near the line of ML. Given the covariance $(\Sigma_0)_{ii} = \alpha\omega_i$, as α increases, the prior becomes less reliable. When α is small, the prior plays an important role in the classification, the predictive solution uses all the information and the prior helps to increases precision of classification.

For the ML solution, the classification error stays unchanged with increasing α. Since the mean and covariance of ML solution are derived from the dataset, they have nothing to do with the parameter α.

For the MAP solution, the classification error increases with the increasing α, and it finally converge very close to the line of ML. The covariance of MAP solution is the same as that of ML, but the mean μ_{MAP} is derived using the same equation as the predictive solution, so MAP uses the prior to some extent. Hence, as α increases, the prior becomes less reliable too, and the classification error increases.

For comparison, under strategy 1, the classification errors are sorted as:

$$\text{Error(Predictive)} < \text{Error(MAP)} < \text{Error(ML)}$$

And the predictive and MAP solution finally converge close to the line of ML when α

is sufficient large. This is because the predictive solution uses all the information, while MAP the prior only to some extent, and ML uses none of the prior. Since the prior is precise, it helps the precision of the solution when α is small, meaning the prior is reliable. And it converges to error somewhat near ML, meaning the reliable information is just the dataset.

In addition, the classification errors of the four datasets are sorted as:

$$\text{Error(D2)} < \text{Error(D4)} < \text{Error(D3)} < \text{Error(D1)}$$

Theoretically, the classification error decreases when the size of the training dataset increases. But considering the size of the largest dataset D4 is still smaller than 1000, the sizes or the four datasets are still not sufficient large, it's reasonable for the result to be like this.

Next, we do all the things above again for the four datasets under strategy 2, and the plots of curves for the four datasets under strategy 2 are shown as below:

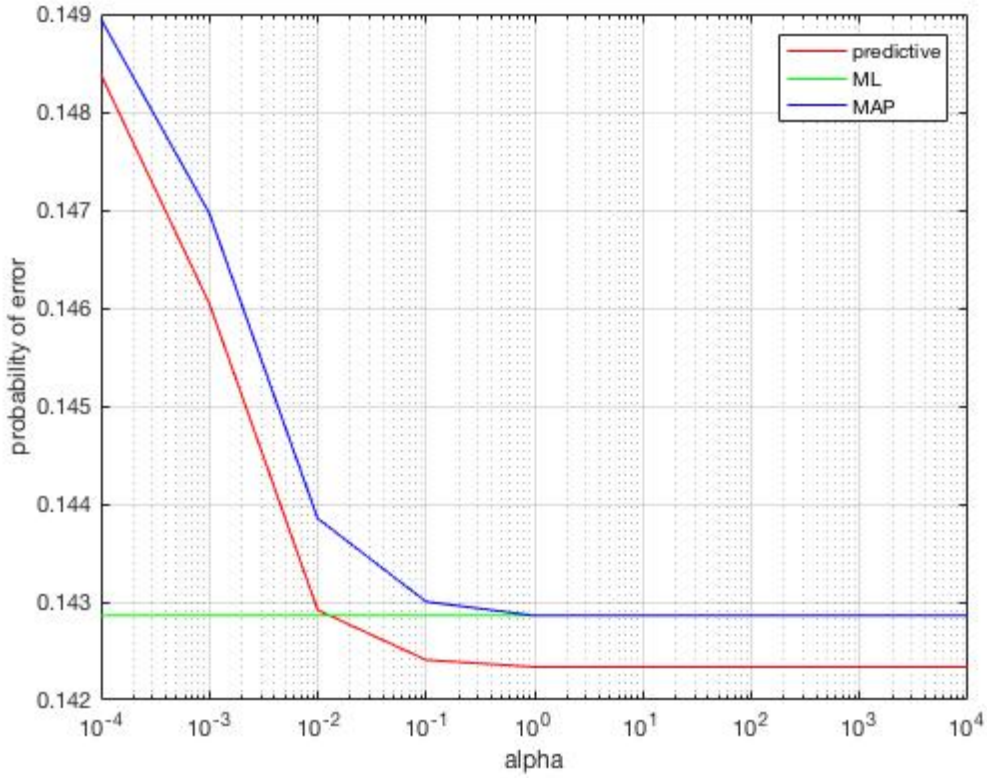

Figure.5 D1 & Strategy 2 Classification Error vs. α

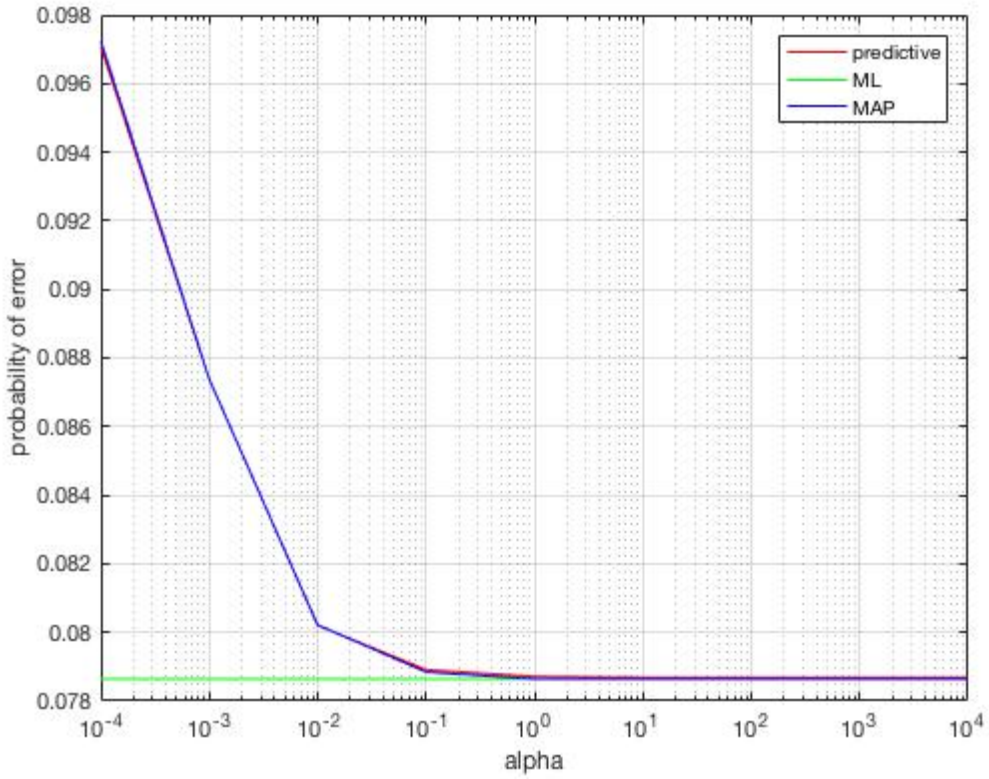

Figure.6 D4 & Strategy 2 Classification Error vs. α

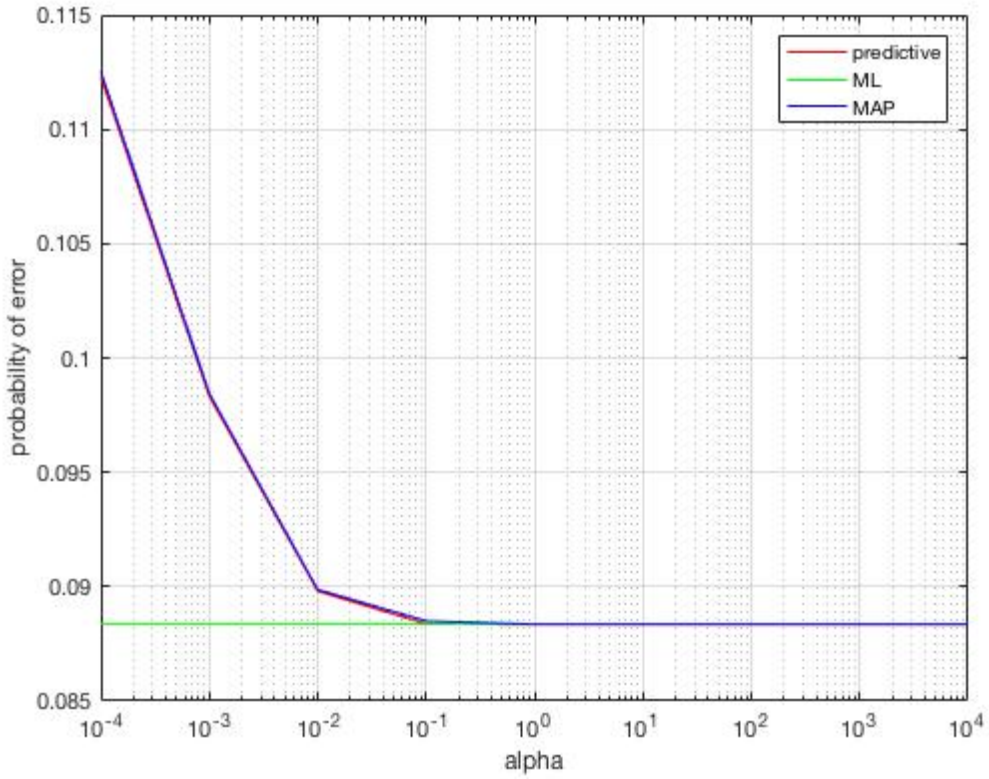

Figure.7 D4 & Strategy 2 Classification Error vs. α

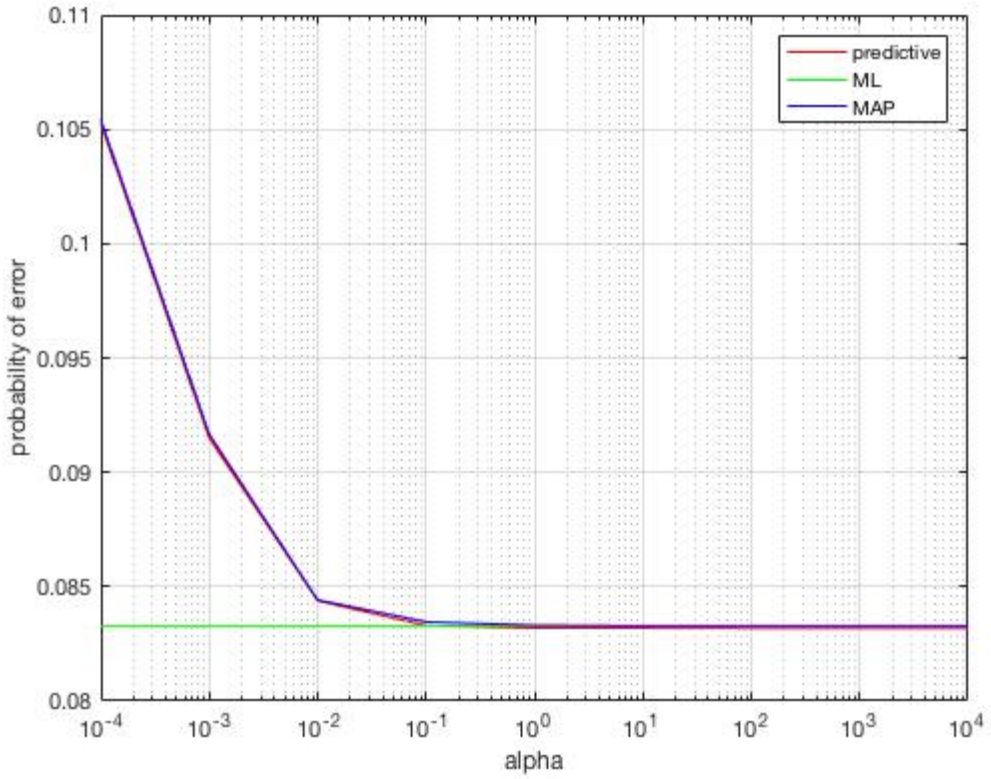

Figure.8 D4 & Strategy 2 Classification Error vs. α

We can see that under strategy 2, for the four plots the classification error of predictive solution decreases with the increasing α, and it finally converge near the line of ML. Since we now consider the prior means of the two classes are the same, the prior is not precise. The predictive solution uses all the information, including the unprecise prior, so the classification error is higher than ML at the beginning. But as mentioned before, when α increases, the prior becomes less reliable, so it finally drops and converge close to the line of ML.

For the ML solution, the classification error stays unchanged with increasing α. And the value of every dataset is the same as those in strategy 1. This is because the mean and covariance of ML solution are derived from the dataset, they have nothing to do with the parameter α and the prior.

For the MAP solution, the classification error also decreases with the increasing α, and it finally converge very close to the line of ML. Similarly, MAP solution uses the prior to some extent, so as α increases, the prior becomes less reliable too.

For comparison, under strategy 2, the classification errors are sorted as:

$$\text{Error(ML)} < \text{Error(Predictive)} < \text{Error(MAP)}$$

And the predictive and MAP solution finally converge close to the line of ML when α

is sufficient large. Unlike strategy 1, now the prior is not precise. So the predictive solution and MAP solution, which use the prior information, now becomes worse than ML. In addition, the MAP is still worse than predictive as expectation. The predictive and MAP solution converges to error somewhat near ML when α is large, because when the prior becomes unreliable, the useful information is just the dataset.

In addition, the classification errors of the four datasets are sorted as:

$$\mathrm{Error(D2)} < \mathrm{Error(D4)} < \mathrm{Error(D3)} < \mathrm{Error(D1)}$$

This is because the size of the four datasets are not sufficient large, so it cannot displace the trend that larger dataset increases the precision.