ITU Computer and Informatics Faculty
BLG 454E Learning From Data, Spring 2018
Homework #2
Due April 10, 2018 11pm

1: Multivariate Analysis

(a) (20 pts) Formulate and implement g(x) discriminant function clearly (add comments) in your code and write its formula into the report.

$$
\begin{aligned}
P(C_i|x) &\propto P(x|C_i) * P(C_i) \\
P(x|C_i) &= \frac{1}{\sqrt{2\pi|\Sigma_i|}} * exp\left(-\frac{1}{2}(x-\mu_i)^T)\Sigma_i^{-}1(x-\mu_i)\right) \\
g_i(x) &= ln(P(C_i|x)) = ln(P(x|C_i) * P(C_i)) \\
g_i(x) &= ln(\frac{1}{\sqrt{2\pi|\Sigma_i|}}) - \frac{1}{2}(x-\mu_i)^T)\Sigma_i^{-}1(x-\mu_i) + ln(P(C_i)) \\
g_i(x) &= -\frac{1}{2}(x-\mu_i)^T)\Sigma_i^{-}1(x-\mu_i) - \frac{1}{2}ln(|\Sigma_i|) - \frac{1}{2}ln(2\pi) + ln(P(C_i))
\end{aligned}
$$

(b) (20 pts) Draw the decision boundaries for each classifier in part(a) for training set is similar to Figure 1 and report it.
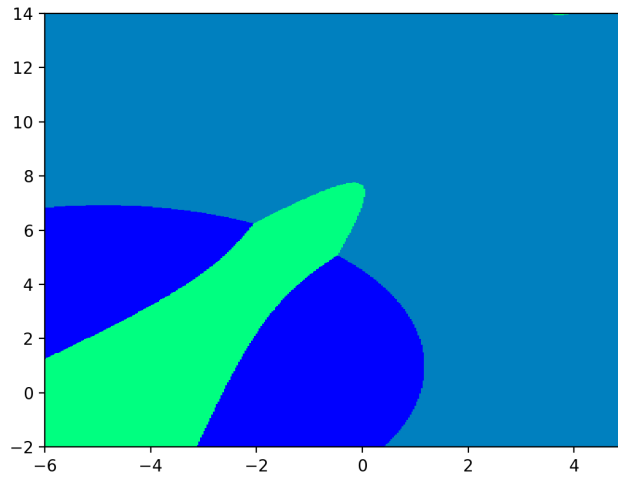

Figure 1: Generated decision boundaries

(c) (10 pts) Calculate test accuracy and write it into the report.

Accuracy: 0.79

1

2: Logistic Regression

(a) (25 pts) Calculate accuracy and confusion matrix using 10 fold cross validation and write them into the report. Which classes are most confused with each other?

Table 1: Accuracy confussion matrix when Learning rate: 0.01, Iteration Count: 250,

	Iris-setosa (Predicted)	Iris-versicolor (Predicted)	Iris-virginica (Predicted)
Iris-setosa (Actual)	50	0	0
Iris-versicolor (Actual)	0	46	4
Iris-virginica (Actual)	0	5	45

Iris-versicolor and Iris-virginica are most confused with each other.

(b) (25 pts) Analyze the effect of learning rate(). Use 10 fold cross validation and compare your results with different learning rates are 10, 1, 0.1, 0.01 based on the number of iterations and classification accuracy.

Learning rate: 0.01, Iteration Count: 250, Accuracy: 0.9466666666666689

Learning rate: 0.1, Iteration Count: 250, Accuracy: 0.9400000000000022

Learning rate: 1, Iteration Count: 250, Accuracy: 0.9466666666666689

Learning rate: 10, Iteration Count: 250, Accuracy: 0.9533333333333356

According to the results, the accuracy of logistic regression is non-linear for learning rates.