Applied Stochastic Processes, 18-751, TX Brown, Fall 2017
Homework #8

Due 8pm Monday October 23.

Read ch 7.1 - 7.6, 8.2 in G&Y. On your own do Quizzes 7.3, 7.4, 7.6, 8.2.

1. Given two classes of data, Bamboo and Mahogany, the back of this sheet contains feature data collected and divided into classes. You should be able to cut and paste the data from the pdf file into your favorite data processing application.

   (a) Plot the two data sets on the same plot using different markers for Bamboo and for Mahogany.

   (b) Compute the following parameters.
   Prior probability: $P[M]$ and $P[B]$
   Mean value: μ_M and μ_B
   Covariance Matrix: K_M and K_B

   (c) Compute two linear estimators, one for Mahogany and one for Bamboo. For each compute the slope and intercept. Draw these lines on your plot in (a).

   (d) Compute the optimal Fisher's Linear Discriminate vector, a. You should do this by hand. Plot the projection vector on the plot in (a).

   (e) Map all the data using a to a single feature denoted, $z = a^T(x, y)$. In z space, plot the two data sets on the same plot using different markers for Bamboo and for Mahogany.

   (f) Compute the following parameters in z space.
   Prior probability: $P[M]$ and $P[B]$
   Mean value: m_M and m_B
   Standard Deviation: σ_M and σ_B
   What is a simple classifier of this data that will be correct 66% of the time.

   (g) Verify that $m_M = a^T \mu_M$, $\sigma_M^2 = a^T K_M a$ and $m_B = a^T \mu_B$, $\sigma_B^2 = a^T K_B a$.

   (h) On the plot in (e) plot $P[z|M]P[M]$ and $P[z|B]P[B]$.

   (i) Compute the MAP decision Boundaries by solving the quadratic equation and indicate the Mahogany and Bamboo decision regions on the graph in (e).

   (j) Estimate the performance of your Fisher's Linear Discriminate classifier in the form P[classifier decision is α, true class is β] in the following table:

		Classifier Decision	
		B	M
True	B	$P[B, B]$	$P[M, B]$
Class	M	$P[B, M]$	$P[M, M]$

   Based on this compute the error rate of this process.
   Note that the performance is an estimate because the parameters are estimated from the data. Note further that even if the data is not precisely a Gaussian distribution, since the projection is a weighted sum of the different dimensions, a Gaussian is a reasonable model for the projection even if the individual input features are not Gaussian.

Mahogany	Bamboo
x y	x y
1.72, 6.05	-2.79, 7.97
-0.19, 5.73	-1.08, 7.06
0.46, 4.51	-1.52, 5.88
-0.27, 5.27	-2.68, 8.09
-1.50, 3.80	-2.16, 10.91
2.99, 4.49	-1.26, 4.01
0.49, 7.38	-1.37, 5.43
1.43, 6.16	-2.51, 8.48
0.29, 3.46	0.01, 4.03
1.54, 3.09	0.41, 3.71
-0.57, 3.64	-0.83, 4.52
2.74, 3.94	-1.66, 5.98
-0.58, 3.84	-1.65, 5.62
0.93, 7.15	1.71, 3.25
1.43, 5.88	-0.87, 6.54
2.05, 5.46	-0.48, 7.75
2.52, 6.19	-0.82, 9.19
0.95, 5.19	-1.86, 8.36
-0.46, 3.90	-2.41, 9.24
0.50, 4.23	-1.66, 7.88
0.51, 4.99	0.01, 4.11
-0.26, 2.39	-1.57, 8.84
2.51, 8.23	-0.31, 5.75
0.86, 5.68	-1.39, 7.29
1.62, 5.40	-1.64, 5.35
2.76, 7.96	-0.21, 9.78
1.99, 5.77	-3.20, 10.92
1.79, 6.29	-3.79, 8.42
0.24, 4.12	-2.33, 5.94
0.73, 6.62	-1.55, 6.13
1.99, 4.47	-1.33, 3.72
0.80, 5.77	-0.50, 6.13
2.81, 8.91	-2.73, 6.71
1.76, 4.26	-2.53, 9.19
-0.40, 4.11	-0.43, 7.18
1.46, 4.65	-1.59, 6.96
0.57, 5.09	0.21, 5.87
2.68, 5.16	-3.44, 5.82
2.29, 6.57	0.06, 7.75
2.45, 5.97	-1.79, 6.23
0.33, 4.10	-3.96, 8.09
0.59, 5.32	-1.94, 6.95
3.31, 5.96	-0.09, 4.76
1.16, 3.70	-2.90, 7.02
0.88, 5.38	-2.25, 11.63
1.04, 5.47	-2.28, 6.67
1.58, 4.76	-2.08, 8.64
0.51, 4.64	-0.62, 5.87
0.81, 4.85	0.19, 1.18
-0.84, 3.38	-2.30, 5.82
2.72, 8.03	
-0.01, 4.90	
2.12, 3.27	
1.58, 6.37	
0.16, 5.37	
0.74, 7.37	
1.58, 5.81	
-0.65, 5.49	
0.12, 6.61	
1.37, 6.17	
0.43, 2.18	
-0.52, 4.66	
0.83, 5.26	
-0.35, 1.92	
2.58, 5.50	
3.09, 6.89	
1.32, 5.78	
0.76, 5.64	
1.93, 7.79	
-0.67, 4.61	
0.84, 8.91	
1.66, 6.72	
0.19, 4.71	
0.31, 3.49	
-0.05, 5.90	
0.82, 5.86	
1.14, 8.16	
1.13, 6.66	
1.54, 5.62	
1.87, 7.50	
0.24, 3.72	
1.59, 4.39	
0.53, 5.57	
1.96, 6.85	
1.01, 5.59	
0.43, 5.53	
-0.07, 2.42	
0.06, 5.20	
0.88, 3.26	
0.82, 6.40	
2.62, 6.99	
0.53, 4.01	
1.03, 5.77	
-0.15, 3.85	
-0.38, 5.41	
-0.54, 5.92	
2.19, 7.59	
0.83, 5.24	
1.76, 7.82	
1.52, 7.59	