MTSC 852 - Pattern Recognition
Introduction to Pattern Recognition
Project Assignments
Due date: Sep. 10, 2015

Sokratis Makrogiannis, Ph.D.

September 4, 2015

Problem 1 (20 points)**.** Pattern Recognition System

1. Describe an application of pattern recognition related to your research. What are the features? What is the decision to be made? Speculate on how one can solve the problem. Limit your answer to one page.

2. Similarly to above item, describe an application of pattern recognition you would be interested in pursuing for fun in your life outside of work.

Problem 2 (30 points)**.** Probability

Show that the normal distribution $N(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ satisfies $E[x] = \int_{-\infty}^{\infty} N(x|\mu, \sigma^2) x dx = \mu$.

Next, by differentiating both sides of $\int_{-\infty}^{\infty} N(x|\mu, \sigma^2) dx = 1$ with respect to σ^2 show that the Gaussian satisfies $E[x^2] = \int_{-\infty}^{\infty} N(x|\mu, \sigma^2) x^2 dx = \mu^2 + \sigma^2$.

Finally show that $var[x] = E[x^2] - E[x]^2 = \sigma^2$ holds true.

Problem 3 (20 points)**.** Diagnostic Test

After your yearly checkup, the doctor has bad news and good news. The bad news is that you tested position for a serious disease, and that the test is 99% accurate. This means that the probability of testing positive given that you have the disease is 0.99, as it is the probability of testing negative given that you do not have the disease. The good news is that this is a rare disease, striking only one in 10,000 people. Why is it good news that the disease is rare? What are the chances that you actually have the disease?

Programming Assignment 1 (30 points)**.** Covariance Matrix

Let $\Sigma_x = \begin{pmatrix} 64 & -25 \\ -25 & 64 \end{pmatrix}$.

1. Verify that Σ_x is a valid covariance matrix.
2. Find the eigenvalues and eigenvectors of Σ_x by hand. Show all your work.
3. Write a program to find and verify the eigenvalues and eigenvectors of Σ_x.
4. We provide 200 data points sampled from distribution $N(0, \Sigma_x)$. Download the accompanying dataset (PR_01_Project.dat) from blackboard and plot the data points. Project the data on the covariance matrix eigenvectors and plot the transformed data. What is the difference between the two plots?

Programming Assignments Write-up For the programming assignments, you are to turn in a brief report (instructions are posted on blackboard page of the course). The report will determine the grade for the programming assignments. Be well organized, type your reports and include figure captions with a short description of all figures in the report. Motivation and initiative are greatly encouraged and will earn extra points.

MTSC 852 - Pattern Recognition
Bayesian Decision Theory
Project Assignments
Due date: Sep. 22, 2015

Sokratis Makrogiannis, Ph.D.

September 16, 2015

Problem 1 (20 points)**.** Normal Distribution

Calculate the entropy of a multidimensional Gaussian $p(\mathbf{x}) = N(\mu, \mathbf{\Sigma})$.

Problem 2 (20 points)**.** Discriminant Functions for Normal Density

Suppose we have two normal distributions with the same covariances but different means: $N(\mu_1, \mathbf{\Sigma})$, $N(\mu_2, \mathbf{\Sigma})$. In terms of their prior probabilities $P(\omega_1)$ and $P(\omega_2)$ state the condition that the Bayes decision boundary not pass between the two means (hint: show that $\mathbf{w}^T(\mu_1 - \mathbf{x}_0)$ and $\mathbf{w}^T(\mu_2 - \mathbf{x}_0)$ have the same sign).

Programming Assignment 1 (30 points)**.** Normal distribution and discriminant functions

1. Write a procedure to generate random samples according to a normal distribution $N(\mu, \mathbf{\Sigma})$ in d dimensions.
2. Write a procedure to calculate the log discriminant function for a given multi-variate Gaussian distribution and prior probability.

3. Write a procedure to calculate the Euclidean distance between two arbitrary points.

4. Write a procedure to calculate the Mahalanobis distance between the mean μ and an arbitrary point $\mathbf{x}$, given the covariance matrix $\mathbf{\Sigma}$.

Apply the above procedures to at least 3 datasets that you will generate in the first step.

Programming Assignment 2 (30 points)**.** Central Limit Theorem

Illustrate the fact that the average of a large number of independent random variables will approximate a Gaussian by the following:

1. Write a program to generate n random integers from a uniform distribution $U(x_l, x_u)$.

2. Write a routine to choose x_l and x_u randomly, in the range $-100 \leq x_l < x_u \leq 100$ and the number of samples n randomly in the range $0 < n \leq 1000$.

3. Generate and plot a histogram of the accumulation of 10^4 points sampled as just described.

4. Calculate the mean and standard deviation of your histogram, and plot it.

5. Repeat the above for 10^5 and 10^6 points. Discuss your results.

Programming Assignments Write-up For the programming assignments, you are to turn in a brief report (instructions are posted on blackboard page of the course). The report will determine the grade for the programming assignments. Be well organized, type your reports and include figure captions with a short description of all figures in the report. Motivation and initiative are greatly encouraged and will earn extra points.

MTSC 852 - Pattern Recognition
Parameter Estimation Techniques
Project Assignments
Due date: Oct. 20, 2015

Sokratis Makrogiannis, Ph.D.

October 13, 2015

Problem 1 (20 points)**.** Problems caused by dimensionality

Here we want to simplify a two-category classifier using shrinkage. Assume two distributions $N(\mu_1, \Sigma_1), N(\mu_2, \Sigma_2)$. Show that one must normalize the data to unit variance in order to be able to shrink an assumed common covariance matrix to identity matrix by applying $\Sigma(\beta) = (1-\beta)\Sigma + \beta\mathbf{I}$, where $0 < \beta < 1$.

Problem 2 (30 points)**.** Fisher Linear Discriminant

Let $p_{\boldsymbol{x}}(\boldsymbol{x}|\omega_i)$ be arbitrary densities with means $\boldsymbol{\mu}_i$ and covariance matrices Σ_i –not necessarily normal– for $i = 1, 2$. Let $y = \boldsymbol{w}^T\boldsymbol{x}$ be a projection, and let the induced one-dimensional densities $p(y|\omega_i)$ have means μ_i and covariances σ_i^2.

1. Show that the criterion function $J_1(\boldsymbol{w}) = \frac{|\mu_1 - \mu_2|^2}{\sigma_1^2 + \sigma_2^2}$ is maximized by $\boldsymbol{w} = (\Sigma_1 + \Sigma_2)^{-1}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})$.

2. If $P(\omega_i)$ is the prior probability for ω_i, show that $J_2(\boldsymbol{w}) = \frac{|\mu_1 - \mu_2|^2}{P(\omega_1)\sigma_1^2 + P(\omega_2)\sigma_2^2}$ is maximized by $\boldsymbol{w} = [P(\omega_1)\Sigma_1 + P(\omega_2)\Sigma_2]^{-1}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})$.

3. Which of these two functions is more closely related to the classic Fisher linear discriminant criterion function $J(\boldsymbol{w}) = \frac{|m_{y1}-m_{y2}|^2}{s_{y1}^2+s_{y2}^2}$? Please explain.

Programming Assignment 1 (50 points)**.** Classification using shrinkage of covariance matrix

Consider the classification of Gaussian data using shrinkage of covariance matrices.

1. Generate 20 training points from each of three equally probable three-dimensional Gaussian distributions $N(\boldsymbol{\mu}_i, Sigma_i)$ with the following parameters

$$\boldsymbol{\mu}_1 = (0,0,0)^T, \quad \Sigma_1 = diag\,[3,5,2]$$

$$\boldsymbol{\mu}_2 = (1,5,-3)^T, \quad \Sigma_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 1 \\ 0 & 1 & 6 \end{pmatrix}$$

$$\boldsymbol{\mu}_3 = (0,0,0)^T, \quad \Sigma_3 = 10\mathbf{I}.$$

2. Write a program to estimate the means and covariances of the data.

3. Write a program that takes α and shrinks these estimated covariance matrices according to $\Sigma_i(\alpha) = \frac{(1-\alpha)n_i\Sigma_i+\alpha n\Sigma}{(1-\alpha)n_i+\alpha n}$, where $0 < \alpha < 1$ is the regularizing parameter, n_i is the number of samples for each class, n is the total number of samples, $i = 1, ..., c$ and c is the number of classes.

4. Plot the training error as a function of α, where $0 < \alpha < 1$.

5. Use your program from part 1 to generate 50 test points from each category. Plot the test error as a function of α.

Programming Assignments Write-up For the programming assignments, you are to turn in a brief report (instructions are posted on blackboard page of the course). The report will determine the grade for the programming assignments. Be well organized, type your reports and include figure captions with a short description of all figures in the report. Motivation and initiative are greatly encouraged and will earn extra points.

MTSC 852 - Pattern Recognition
Nonparametric Estimation Techniques
Project Assignments
Due date: November 12, 2015

Sokratis Makrogiannis, Ph.D.

November 5, 2015

Problem 1 (30 points). Properties of Metrics

Prove that the Minkowski metric indeed posseses the four properties required of all metrics.

Programming Assignment 1 (35 points). Density estimation

Explore some of the properties of density estimation in the following way.

1. Write a program to generate points according to a uniform distribution in a unit cube, $-1/2 \leq x_i \leq 1/2$ for $i = 1, 2, 3$. Generate 104 such points.

2. Write a program to estimate the density at the origin based on your 104 points as a function of the size of a cubical window function of size h. Plot your estimate as a function of h, for $0 < h \leq 1$.

3. Evaluate the density at the origin using n of your points and the volume of a cube window which just encloses n points. Plot your estimate as a function of $n = 1, ..., 104$.

4. Write a program to generate 104 points from a spherical Gaussian density centered on the origin. Repeat (2) & (3) with your Gaussian data.

5. Discuss any qualitative differences between the functional dependencies of your estimation results for the uniform and Gaussian densities.

	ω_1			ω_2			ω_3		
sample	x_1	x_2	x_3	x_1	x_2	x_3	x_1	x_2	x_3
1	0.28	1.31	-6.2	0.011	1.03	-0.21	1.36	2.17	0.14
2	0.07	0.58	-0.78	1.27	1.28	0.08	1.41	1.45	-0.38
3	1.54	2.01	-1.63	0.13	3.12	0.16	1.22	0.99	0.69
4	-0.44	1.18	-4.32	-0.21	1.23	-0.11	2.46	2.19	1.31
5	-0.81	0.21	5.73	-2.18	1.39	-0.19	0.68	0.79	0.87
6	1.52	3.16	2.77	0.34	1.96	-0.16	2.51	3.22	1.35
7	2.20	2.42	-0.19	-1.38	0.94	0.45	0.60	2.44	0.92
8	0.91	1.94	6.21	-0.12	0.82	0.17	0.64	0.13	0.97
9	0.65	1.93	4.38	-1.44	2.31	0.14	0.85	0.58	0.99
10	-0.26	0.82	-0.96	0.26	1.94	0.08	0.66	0.51	0.88

Table 1: Three-dimensional data sampled from three categories.

Programming Assignment 2 (35 points)**.** Consider k-nearest-neighbor density estimations in different numbers of dimensions

1. Write a program to find the k-nearest-neighbor density for n (unordered) points in one dimension. Use your program to plot such a density estimate for the x_1 values in category ω_3 in the table above (and in file "ch4_dhs_samples.dat") for $k = 1$, 3 and 5.

2. Write a program to find the k-nearest-neighbor density estimate for n points in two dimensions. Use your program to plot such a density estimate for the $(x_1, x_2)^T$ values in ω_2 for $k = 1$, 3 and 5.

3. Write a program to form a k-nearest-neighbor classifier for the three-dimensional data from the three categories in the table above. Use your program with $k = 1, 3, 5$ to estimate the relative densities at the following points: $(-0.41, 0.82, 0.88)^T$, $(0.14, 0.72, 4.1)^T$ and $(-0.81, 0.61, -0.38)^T$.

Programming Assignments Write-up For the programming assignments, you are to turn in a brief report (instructions are posted on blackboard page of the course). The report will determine the grade for the programming assignments. Be well organized, type your reports and include figure captions with a short description of all figures in the report. Motivation and initiative are greatly encouraged and will earn extra points.

MTSC 852 - Pattern Recognition
Linear Discriminant Functions
Project Assignments
Due date: November 30, 2015

Sokratis Makrogiannis, Ph.D.

November 23, 2015

Problem 1 (30 points)**.** Linear discriminant function properties

Consider a linear machine with discriminant functions $g_i(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x} + w_{i0}$, $1, \ldots, c$. Show that the decision regions are convex by showing that if $\boldsymbol{x}_1 \in \mathcal{R}_i$ and $\boldsymbol{x}_2 \in \mathcal{R}_i$, then $\lambda\boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 \in \mathcal{R}_i$ for $0 \leq \lambda \leq 1$.

Problem 2 (30 points)**.** Gradient descent

Derive the equation $n(k) = \frac{\|\boldsymbol{\nabla} J\|}{\boldsymbol{\nabla} J^T H \boldsymbol{\nabla} J^T}$, where $J(.)$ depends on iteration step k for the gradient descent algorithm.

Programming Assignment 1 (40 points)**.** Batch perceptron

Write a program to implement the batch Perceptron algorithm.

1. Starting with $\boldsymbol{a} = \mathbf{0}$, apply your program to the training data from ω_1 and ω_2. Note the number of iterations required for convergence.
2. Apply your program to ω_3 and ω_2. Again note the number of iterations required for convergence.
3. Explain the difference between the iterations required in the two cases.

sample	ω_1		ω_2		ω_3		ω_4	
	x_1	x_2	x_1	x_2	x_1	x_2	x_1	x_2
1	0.1	1.1	7.1	4.2	-3.0	-2.9	-2.0	-8.4
2	6.8	7.1	-1.4	-4.3	0.5	8.7	-8.9	0.2
3	-3.5	-4.1	4.5	0.0	2.9	2.1	-4.2	-7.7
4	2.0	2.7	6.3	1.6	-0.1	5.2	-8.5	-3.2
5	4.1	2.8	4.2	1.9	-4.0	2.2	-6.7	-4.0
6	3.1	5.0	1.4	-3.2	-1.3	3.7	-0.5	-9.2
7	-0.8	-1.3	2.4	-4.0	-3.4	6.2	-5.3	-6.7
8	0.9	1.2	2.5	-6.1	-4.1	3.4	-8.7	-6.4
9	5.0	6.4	8.4	3.7	-5.1	1.6	-7.1	-9.7
10	3.9	4.0	4.1	-2.2	1.9	5.1	-8.0	-6.3

Table 1: Two-dimensional data sampled from three categories.

Programming Assignments Write-up For the programming assignments, you are to turn in a brief report (instructions are posted on blackboard page of the course). The report will determine the grade for the programming assignments. Be well organized, type your reports and include figure captions with a short description of all figures in the report. Motivation and initiative are greatly encouraged and will earn extra points.