CSCI 5521: Pattern Recognition

Please submit your work in an electronic document with a standard readable format (e.g. pdf, rtf, doc, txt). All matlab code should be put into a separate file that is executable as a script or function.

Problem set 1

Due: 9/20/07 Midnight

Download the file arrow.m:
http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=278&objectType=file
Use the arrow command to visualize vectors.

1. (10%) Consider the vectors $x_1 = [1,0]^T$; $x_2 = \left[1,\frac{1}{\sqrt{2}}\right]^T$; Compute the area formed by the parallelogram of these two vectors using the formula for the area of a right triangle $Area = \frac{1}{2}(base * height)$. Now put the two vectors into a matrix **A**, and compute the determinant. What is the relationship between det(**A**) and area? What if either one or both x_i is negated? Now let $x_1 = [0.86603,0.5]^T$; $x_2 = [0.51247,1.11237]^T$; Again use the right triangle formula and the determinant to determine the area (it may help to visualize using arrow.m) and the det(). Can you find a relationship between these two formulae?

2. (20%) Consider the equation $\vec{y} = A\vec{x} \Rightarrow \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} b_1^T \\ b_2^T \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

   Let

   $b_1 = \begin{bmatrix} -\sqrt{3}/4 \\ \frac{1}{4} \end{bmatrix}; \quad b_2 = \begin{bmatrix} \sqrt{3}/4 \\ \frac{1}{4} \end{bmatrix};$

   Using Matlab, plot the constraint lines determined by b_1 and b_2 for the range –4:4, i.e. determine the lines perpendicular to b_1 and b_2 such that the dot product between the b_i and any other vector is equal to {-4,-3,-2,…,4} for each b_i, and display them using the line() command or the plot() command. Graphically solve for x when $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$. Use matrix inversion to solve for x.

3. (20%) Let x and y be identically Gaussian random variables:

   $p(x) = N(\mu,\sigma_x) = \frac{1}{\sqrt{2\pi}\sigma_x}\exp(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma_x^2})$

   $p(y \mid x) = N(x,\sigma_y) = \frac{1}{\sqrt{2\pi}\sigma_y}\exp(-\frac{1}{2}\frac{(y-x)^2}{\sigma_y^2})$

Choose values for μ & σ. Evaluate *p(x)* and *p(y | x)* over a grid of points *x* and *y* over appropriate ranges (so that *p*(*x*) drops close to zero"within the grid). The matlab command meshgrid() may help. Treat the resulting probability and grid values as a discrete probability distribution, and construct probability tables for both distributions. The purpose is to approximate the continuous distributions by discrete distributions. Make sure the sum-to-1 constraints are satisfied. Using the discrete tables, use marginalization and Bayes' rule to compute the probability tables for *p(y)* and *p(x|y)*.

4. (15%) Generate a set of 100 random *x,y* coordinates using rand(). Use scatter() to display them. All the points should live in the unit square. Let *s* denote 1 point. Compute the average inner product of these points: $E[s^T \cdot s] = \frac{1}{100} \sum_{j=1:100} s_j^T \cdot s_j$.

Transform each of these points by a matrix $Q = \begin{bmatrix} 0.7419 & 4.2243 \\ 0.7283 & -1.3573 \end{bmatrix}$, so that $z = Q \cdot s$. Scatterplot z. Compute the average inner product of the *z* points. Use an eigenanalysis of Q to explain the difference between the *s* and the *z* lengths. Compute the area the transformed points live in using Q. Using the relationship $Q=A\Lambda A^{-1}$, (*where A* contains the normalized eigenvectors and Λ is a diagonal matrix of eigenvalues) to determine the transform executed by *A*. Use arrow() to plot the columns of A with the scatterplot of *z*. The net effect of *Q* on the length of the vectors *s* is a rotation and a scaling introduced by Q^TQ. Find the angle of the rotation executed by Q^TQ. Hint: do an eigenanalysis of Q^TQ, make sure the eigenvectors (e.g. *A*) are length 1 and the eigenvector matrix has det of 1 (the det = -1, so you must fix it), then use the parametric form of a 2-D rotation matrix:

$$R = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

5. (15%) Compute the distribution of the sum of four fair 8-sided dice.

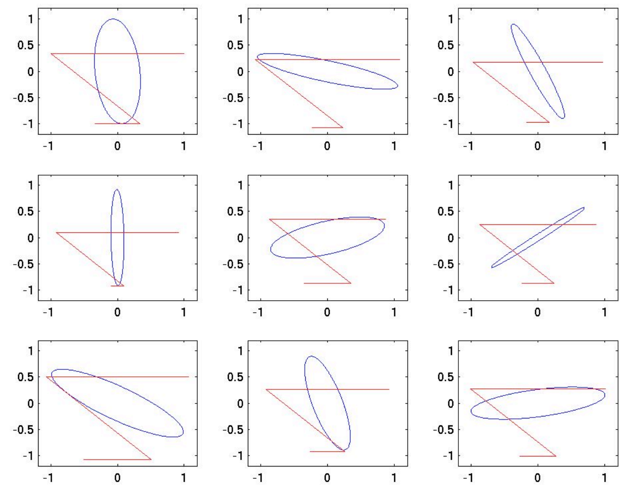
6. (20%) Below you will find 2 sets (A and B), each with 9 examples of patterns generated by some mysterious process, labeled training data. Below that you will find an additional set labeled test data.

TRAIN - Using only the training data, invent two different measurements on aspects of these patterns, and make these measurements on all the patterns (by a method of your choice – for example by hand with a ruler, via a graphics program, etc.). Try to choose measurements that help separate the classes. Call the first measurement type feature1, and the second feature2. Make a scatter plot of the feature1 and feature2 values, using plot or scatter. Use different labels for class 1 and class 2. Determine a straight line in this space that appears to separate this data by eye.

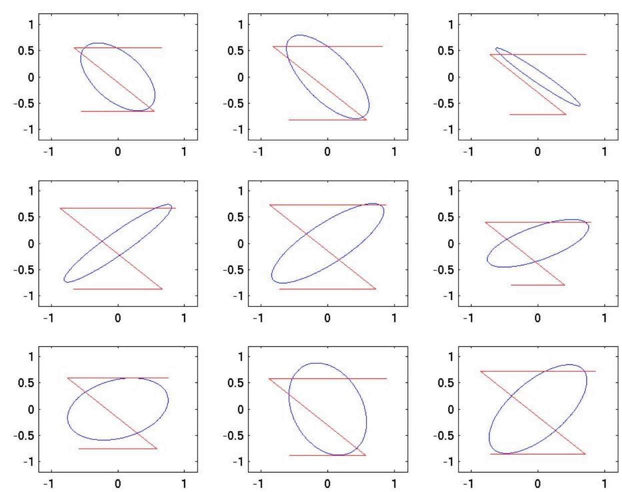
Now determine the measurements on the TEST DATA, and plot the new feature vectors *in the same plot (i.e. using the matlab command hold on), but using symbols that distinguish both class A and B and TEST points from TRAINING points.* Determine the number of errors and the type of error (e.g. labeled A when actually B) for the TEST DATA using your line to classify. Submit the resulting image.

TRAINING DATA

SET A



SET B



TEST DATA