CS 395T # Homework 3 **Due: Mar 21, 2007**

Use MATLAB for question 2, and hand in the code along with your results. Note that the assignment is due IN CLASS.

1. The *multinomial* distribution is a multi-dimensional generalization of the binomial distribution. Given a parameter vector $\mathbf{a} = [a_1, ..., a_p]$, with $\sum_i a_i = 1$ and $a_i > 0$, and a positive integer n, the probability density function over $\mathbf{x} = [x_1, ..., x_p]$ (where each x_i is an integer) is given by

$$p(\mathbf{x}) = \frac{n!}{x_1!...x_p!} \prod_{i=1}^{p} a_i^{x_i},$$

when $x_i \geq 0$ and $\sum_{i=1}^{p} x_i = n$ ($p(\mathbf{x}) = 0$ otherwise).

a) Given data from two classes C_1 and C_2, assume that $p(\mathbf{x}|C_j)$ is described by a multinomial distribution with parameters $\mathbf{a}^j$ and n, and consider classification of a point $\mathbf{x}$. Using Bayes decision theory, show that this leads to a linear decision boundary for classification.

b) Now consider k classes, and assume further that $p(C_j)$ is the same for all classes. Show that finding the class j to maximize $p(C_j|\mathbf{x})$, i.e. finding $\text{argmax}_j p(C_j|\mathbf{x})$, can equivalently be expressed as the following optimization problem:

$$\text{argmin}_j KL(\tilde{\mathbf{x}}, \mathbf{a}^j),$$

where $\tilde{\mathbf{x}} = \mathbf{x}/\|\mathbf{x}\|_1$ and $KL(\tilde{\mathbf{x}}, \mathbf{a}^j)$ denotes the KL-divergence between $\tilde{\mathbf{x}}$ and $\mathbf{a}^j$. So for the multinomial distribution, the decision rule for classification of a point involves finding the closest parameter vector in KL-divergence; note the similarity to the case when the data comes from spherical Gaussians, where the squared Euclidean distance to the means is used instead. What happens when $p(C_j)$ is not the same for every class?

2. Download the data set `www.cs.utexas.edu/users/kulis/dm2007/3gauss.mat` using the command `load 3gauss`. The training data is `X`, the training labels `X_lab`, the test data `X_test`, and the test labels `X_lab_test`. Form the between-class S_B and within-class S_W scatter matrices of the training data for Fisher's linear discriminant analysis. Then perform, discuss, and compare the following experiments:

a) Classify the test data by assigning the test points to the nearest class mean of the training set.

b) Classify the test data by projecting the test data onto the leading eigenvector of the Fisher linear discriminant projection and assigning the points to the nearest class means of the projected training points.

c) Classify the test data by projecting the test data onto the leading two eigenvectors of the Fisher linear discriminant projection and assigning the points to the nearest class means of the projected training points.

d) Repeat this experiment on the iris data set. You can download the iris data set from `www.cs.utexas.edu/users/kulis/dm2007/iris_data.mat`.