CS6670 Assignment 3
Target due date: April 23, 2015

The homework is generally split into programming exercises and written exercises. Target due dates will be given for each homework, however you can submit all your homework, without penalty, at the end of the semester. You are encouraged to keep up with the target due dates. You should turn in an electronic copy of your solutions to the homework. Please submit your homework on the course website. You are responsible for submitting clear, organized answers to the questions. Please include all relevant information for a question, including text response, equations, figures, graphs, etc. Please pay attention to the discussion board for relevant information regarding updates, tips, and policy changes.

1 WRITTEN EXERCISES

1. Mahalanobis distance.
   The Mahalanobis distance between \( x^i \) and \( x^j \) is given by \( \Delta^2 = (x^i - x^j)^\top \Sigma^{-1} (x^i - x^j) \), where \( \Sigma \) is a \( d \times d \) covariance matrix.
   (a) A covariance matrix \( \Sigma \), by definition, is symmetric and positive definite, which means \( a^\top \Sigma a > 0 \) for all \( a \in \mathbb{R}^d \). Show that a necessary and sufficient condition for \( \Sigma \) to be positive definite is that all of its eigenvalues are positive.
   (b) \( \Delta^2 \) is equivalent to the squared Euclidean distance between \( y^i \) and \( y^j \), where \( y \) is a linearly transformed version of \( x \). What is that transformation?
   (c) Give an example of an application for which Mahalanobis distance is appropriate (e.g., compared to \( L_2 \) distance) and explain intuitively what \( \Sigma^{-1} \) captures in this case.

2. Properties of Chi Squared distance.
   Recall that the \( \chi^2 \) distance is given by \( \chi^2_{ij} = \frac{1}{2} \sum_{k=1}^{d} \frac{(x^i_k - x^j_k)^2}{(x^i_k + x^j_k)} \) where the \( x \)'s are normalized histogram vectors. Prove or disprove the following statements:
   (a) \( \chi^2_{ij} \in [0, 1] \).
   (b) The matrix \( Q \in \mathbb{R}^{n \times n} \) with entries \( Q_{ij} = \sum_{k=1}^{d} \frac{x^i_k x^j_k}{(x^i_k + x^j_k)} \) is positive definite.
   (c) \( \chi^2_{ij} \) is a metric.
2 Programming Exercises

1. Tiny Images.

(a) Download the CIFAR-10 training and testing data from http://www.cs.toronto.edu/kriz/cifar.html.

(b) Write a utility to extract the images (of size $32 \times 32 \times 3$) and labels (0, … , 9). Use it to import the first $M = 2000$ training images and the first $N = 1000$ testing images.

(c) Display the first 40 training images together with their labels, arranged in a $4 \times 10$ array.

(d) Compute the prior probability of each class in the training set. Is it uniform?


Convert the images into grayscale and let $x^i \in \mathbb{R}^d$ (with $d = 32^2$) denote the $i$th training example concatenated as a column vector.

(a) Implement the following pairwise comparison functions of the form $\mathcal{D}(x^i, x^j)$:

- $L_p$ norm: $\left( \sum_{k=1}^{d} |x^i_k - x^j_k|^p \right)^{1/p}$
- Inner product: $(x^i)^\top x^j$
- Normalized inner product: $(x^i)^\top x^j / \|x^i\| \|x^j\|$
- $\chi^2$ distance: $\frac{1}{2} \sum_{k=1}^{d} \frac{(x^i_k - x^j_k)^2}{(x^i_k + x^j_k)}$

Each is defined for $x \in \mathbb{R}^d$ except $\chi^2$, which requires $x$ to be nonnegative and sum to 1.

(b) Compute and display the best match (using max or min as appropriate) for the first 10 training images (excluding self matches) vs. all $M$ training images using $L_1$, $L_2$, $L_\infty$, and inner product (both normalized and raw). Use an asterisk to indicate errors.

(c) Which choice of $\mathcal{D}(\cdot, \cdot)$ gave the fewest errors? Which gave the most?

3. Confusion Matrices and ROC Curves.

(a) Compute the $L_2$ distance from all $N$ testing images to all $M$ training images.

(b) Assuming a 1-nearest neighbor classifier, compute the $10 \times 10$ confusion matrix for this experiment. Display it as an image and comment on what it reveals about the classification behavior for classes such as “bird” and “cat”.

(c) Compute the histogram of distances for genuine matches and for impostors. Plot the two histograms on the same set of axes.

(d) Plot the ROC curve for this experiment. What is the equal error rate?

4. Dimensionality reduction on the CIFAR-10 database.

(a) Generate a random projection matrix $G \in \mathbb{R}^{d' \times d}$ with entries $G_{ij} \sim \mathcal{N}(0, 1/d')$. Use $d' = 64$, which represents a factor of 16 smaller than the full dimensionality ($d = 1024$). Compute the mean squared difference between the entries of $G^\top G$ and a $d \times d$ identity matrix. It should be close to $1/d'$. 

2
(b) Compute the ROC curve as in problem 3 using $L_2$ distances on $Gx^j$ in place of $x^j$. How does the new EER compare to the old one?

(c) Repeat the preceding step with a different dimensionality reduction method of your choice, keeping $d'$ fixed.