Homework Set Two
ECE 271A
Department of Computer and Electrical Engineering
University of California, San Diego

This HW set contains several problems. Only the problem labeled **Quiz** must be handed in and will be graded. The remaining problems are for practice. You should not submit them for grade. By submitting your Quiz solution, you agree to comply with the following.

1. The Quiz is treated as a **take-home test** and is an **INDIVIDUAL** effort. **NO collaboration is allowed**. The submitted work must be yours and must be original.
2. The work that you turn-in is your own, using the resources that are available to **all** students in the class.
3. You can use the help of **GENERAL** resources on programming, such as MATLAB tutorials, or related activities.
4. You are not allowed to consult or use resources provided by tutors, previous students in the class, or any websites that provide solutions or help in solving assignments and exams.
5. You will not upload your solutions or any other course materials to any web-sites or in some other way distribute them outside the class.
6. 0 points will be assigned if your work seems to violate these rules and, if recurrent, the incident(s) will be reported to the Academic Integrity Office.

1. Problem 2.6.26 in Duda, Hart, and Stork (DHS).

2. In this problem we will consider the ML estimate of the parameters of a multinomial distribution. Consider a random variable X such that $P_X(k) = \pi_k, k \in \{1, \ldots, N\}$. Suppose we draw n independent observations from X and form a random vector $\mathbf{C} = (C_1, \ldots, C_N)^T$ where C_k is the number of times that the observed value is k (i.e. $\mathbf{C}$ is the histogram of the sample of observations). Then, $\mathbf{C}$ has multinomial distribution

$$P_{C_1,\ldots,C_N}(c_1, \ldots, c_N) = \frac{n!}{\prod_{k=1}^{N} c_k!} \prod_{j=1}^{N} \pi_j^{c_j}.$$

a) Derive the ML estimator for the parameters $\pi_i, i = 1, \ldots, N$. (Hint: notice that these parameters are probabilities, which makes this an optimization problem with a constraint. If you know about Lagrange multipliers feel free to use them. Otherwise, note that minimizing a function $f(a, b)$ under the constraint $a + b = 1$ is the same as minimizing the function $f(a, 1 - a)$).

b) Is the estimator derived in **a)** unbiased? What is its variance? Is this a good estimator? Why?

3. Problem 3.2.8 in DHS.

4. Problem 3.2.10 in DHS. Assume that the random variables $X_1, \ldots, X_n$ are iid with a distribution of mean μ, which is the quantity to estimate.

5. In this problem we will consider the ML estimate of the Gaussian covariance matrix.

a) Problem 3.4.13 in DHS.

b) Derive the same result by computing derivatives in the usual way. (Hint: you may want to use a manual of matrix calculus such as that at `http://www.ee.ic.ac.uk/hp/staff/dmb/matrix/calculus.html`. Also, it may be easier to work with the precision matrix $\mathbf{P} = \mathbf{\Sigma^{-1}}$.)

6. (Quiz) This week we will continue trying to classify our cheetah example. Once again we use the decomposition into 8×8 image blocks, compute the DCT of each block, and zig-zag scan. However, we are going to assume that the class-conditional densities are multivariate Gaussians of 64 dimensions.

Note: The training examples we used last time contained the absolute value of the DCT coefficients instead of the coefficients themselves. Please download the file `TrainingSamplesDCT_8_new.mat` and use it in this and all future exercises. For simplicity, I will still refer to it as `TrainingSamplesDCT_8.mat`.

a) Using the training data in `TrainingSamplesDCT_8.mat` compute the histogram estimate of the prior $P_Y(i), i \in \{cheetah, grass\}$. Using the results of problem **2** compute the maximum likelihood estimate for the prior probabilities. Compare the result with the estimates that you obtained last week. If they are the same, interpret what you did last week. If they are different, explain the differences.

b) Using the training data in `TrainingSamplesDCT_8.mat`, compute the maximum likelihood estimates for the parameters of the class conditional densities $P_{X|Y}(x|cheetah)$ and $P_{X|Y}(x|grass)$ under the Gaussian assumption. Denoting by $\mathbf{X} = \{X_1, \ldots, X_{64}\}$ the vector of DCT coefficients, create 64 plots with the marginal densities for the two classes - $P_{X_k|Y}(x_k|cheetah)$ and $P_{X_k|Y}(x_k|grass), k = 1, \ldots, 64$ - on each. Use different line styles for each marginal. Select, by visual inspection, what you think are the best 8 features for classification purposes and what you think are the worst 8 features (you can use the `subplot` command to compare several plots at a time). Hand in the plots of the marginal densities for the best-8 and worst-8 features (once again you can use `subplot`, this should not require more than two sheets of paper). In each subplot indicate the feature that it refers to.

c) Compute the Bayesian decision rule and classify the locations of the cheetah image using i) the 64-dimensional Gaussians, and ii) the 8-dimensional Gaussians associated with the best 8 features. For the two cases, plot the classification masks and compute the probability of error by comparing with `cheetah_mask.bmp`. Can you explain the results?