Homework Set Four

ECE 175

Department of Electrical and Computer Engineering
University of California, San Diego
Nuno Vasconcelos

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. Consider the binomial random variable X with parameters n and p, i.e.

$$P_X(x) = \left(\begin{array}{c} n \\ x \end{array} \right) p^x (1-p)^{n-x}$$

Assuming that the parameter n is known, and given a sample $\mathcal{D} = \{x_1, \ldots, x_N\}$, what is the the maximum likelihood estimate of the parameter p?

2. Consider a d-dimensional random variable

$$\mathbf{Y} = \mathbf{A}\mathbf{x} + \mathbf{n}$$

where $\mathbf{x}$ is an unknown, but deterministic, d-dimensional vector $\mathbf{n}$ a d dimensional Gaussian random vector of mean $\mathbf{0}$ and covariance $\mathbf{\Sigma}$ and $\mathbf{A}$ a $n \times n$ positive definite matrix.

a) what is the joint density for the random variable $\mathbf{Y}$?

b) show that, given an observation $\mathbf{y}$, the maximum likelihood estimate of $\mathbf{x}$ is

$$\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{\Sigma}^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{\Sigma}^{-1} \mathbf{y}.$$

c) what is the least squares problem whose solution is equivalent to **b)**? Assume $\mathbf{\Sigma}$ is diagonal. What is the role of this matrix, i.e. how does it change the canonical least squares problem?

Quiz (Computer) In last two problems, we saw how classification of hand written digits works. This time we will move to a more general and practical situation. The digital scan of the digits as in Fig.1 is often noisy and may vary a lot in terms of overall intensity. We want to classify these corrupted digits. We shall continue with the training data used in the previous experiment, but with a new set of test data `testImagesNew`, which is 1) corrupted by noise and 2) re-scaled in amplitude. Using the Nearest Neighbor approach, we will find the training image that is nearest to the test image, but instead of the Euclidean distance, we will use the distance to the ML estimate of the uncorrupted test pattern. We assume the test image $\mathbf{Y}$ is of the form

$$\mathbf{Y} = a * \mathbf{x} + \mathbf{N}$$

i.e. the result of corrupting the training image image $\mathbf{x}$ through 1) amplitude re-scaling by the scalar a, and 2) addition of independent zero mean Gaussian noise with variance v, i.e. $\mathbf{N} \sim \mathcal{G}(\mathbf{0}, v\mathbf{I})$. The scale factor a is the maximum likelihood estimate given the observation $\mathbf{y}$ (the test image) and the known training image $\mathbf{x}$. The Euclidean distance between the normalized test image and the training image then serves as the metric for the NN classifier.

$$a^* = argmax_a \;\; P_Y(Y|X, a, v)$$

1. Using the two sample images `sampletest.png` and `sampletrain.png` (i) calculate the ML estimate of the scale parameter 'a'.

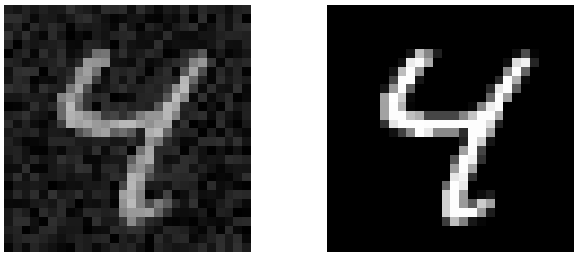

Figure 1: Sample Images

2. Now for the new testset `testImagesNew`, perform the task of classification using the least square distance metric. As before, Compute and plot the error rates for each class, and the total error rate.

3. Perform a NN classification on the new testset, using the algorithm of Computer Problem 1, i.e. using Euclidean distance metric. Compare your results with the NN classification performed in part 2.

   Notes:

   - Remember to convert everything to double before calculating the ML estimate of a.
   - All the data needed for this assignment is uploaded on the website do not use data from the previous experiments.