HW #5 Due: 5/16/2022

1. If we want to use the approach given on the lecture notes to perform dimensionality reduction based on ICA, can we directly pick independent components with larger energy? Explain.
2. Suppose we have two classes of data. The samples in the first class are from $\begin{bmatrix} \mathrm{x}_1 \\ \mathrm{x}_2 \end{bmatrix}$ where x_1 and x_2 are jointly Gaussian random variables (r.v.) with $\mu_\mathrm{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $\Sigma_\mathrm{x} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$. The samples in the second class are from $\begin{bmatrix} \mathrm{y}_1 \\ \mathrm{y}_2 \end{bmatrix}$ where y_1 and y_2 are jointly Gaussian r.v. with $\mu_\mathrm{y} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$ and $\Sigma_\mathrm{y} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$. Find the $\boldsymbol{w}_{opt}$ (with normalization) and optimal decision value α_{opt} based on the concept of the linear discriminant analysis (LDA).
3. Use the Lagrange multiplier method to find the solution of the following: Minimize $f(x, y) = x^2 + y^2$ subject to $x + y = 2$.
4. Use the PCA approach to reduce the feature dimension of the Iris data set from 4 to 3. As usual, take 70% of the samples as training set and perform PCA. Use 5-NN to classify the test set and then report the average accuracy after 10 trials.
5. We mention the gradient descent algorithm in the lecture. You are asked to write a gradient descent program to find the minimum of the following cost function
$$J(x, y) = x + y - 100(x^2 + y^2 - 1)^2.$$
To have a unique answer, use $x_0 = y_0 = 1$ and $\eta = 0.005$. (a) Iterate the program 1,000 times and print out the final values of (x, y). (b) Is your answer close to the result obtained from the Lagrange multiplier (the example in the lecture notes)? (c) If the analytic form of the $J(\mathrm{x}, \mathrm{y})$ is unknown and only input/output pairs are available (like a black box), how do you estimate the derivatives to complete the computation?