HW #3 Due: 5/02/2019

1. Show that the (excess) kertosis of a Gaussian random variable is zero.
2. In this problem, you are asked to use the LDA to reduce the dimensionality of the breast cancer dataset before classification. (a) If only “benign” or “malignant” is to be determined, what is the highest dimension of features after LDA reduction? (b) With your answer in (a), write an LDA program and find the average accuracy with 3-NN. To begin one experiment, randomly draw 70 % of the instances from each class for training and the rest for testing. Repeat the drawing and the k-NN classification 10 times and compute the average accuracy. Use average values to fill the missing attributes.
3. We learned the k-means for clustering in the lecture. Implement the algorithm with the breast cancer dataset. In this problem, we know $k = 2$. Use the first two samples in the dataset as the initial cluster center to do the clustering. Remember that the cluster centers are points in 9-dimensional space. To have a unique answer, use the same sequence of samples given in the dataset to feed into your program. That is, do not shuffler the dataset. Once your program converges, (a) print out the coordinates of the cluster centers, (b) and the number of members (sample points) in each cluster. (c) According to the labels of data samples, how many of them are placed in wrong clusters? Use a majority vote to determine the label of each cluster.
4. We used the play/no play example in the lecture. You are required to write a program to build an ID3 decision tree for the breast cancer dataset. In this dataset, we have nine attributes and each attribute has at most 10 distinct values. Build such a tree and test the accuracy. Again, remember the 70/30 split in training and testing and repeat experiments 10 times. Use average values (with rounding) to fill the missing attributes.
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? (c) If we want to make the program a general-purpose one, we have to assume that $\mathrm{J}(\mathrm{x}, \mathrm{y})$ is a black box (only input/output pairs are available without explicit expression). If that is the case, how do you estimate the derivatives to complete the computation?