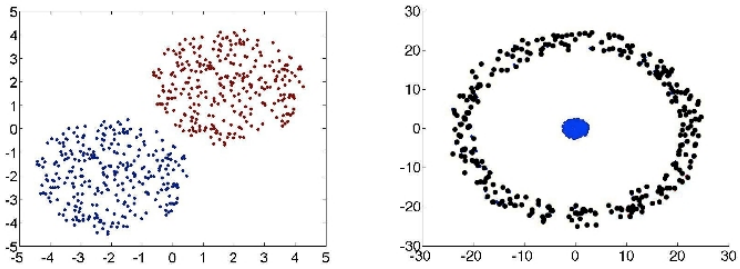
CS 380 Machine Learning - Spring 2011
Homework Assignment 4
Due at the start of class on March 30^{th}

1. (Bias/Variance – 25pts) For this question you will experiment with bagging and boosting. These algorithms are available in Weka under the Meta category. In both cases, use the J48 decision tree algorithm as the base classifier and set numIterations to 30. Set the J48 unpruned option to true. For boosting, set the weightThreshold to 1000.

   To use J48, after choosing bagging (or boosting) from the Meta category, click on the text that starts Bagging -P 100 beside the Choose button. Then click the Choose button beside classifier on the window that pops up and select J48. You should now see J48 -C 0.25 -M 2 beside the Choose button. Now click that text to set the unpruned option.

   Choose a dataset provided by Weka (one that has more than 100 instances) and compare the performance of J48, bagging, and boosting in terms of classification accuracy. What are the effects of bagging and boosting, and how can they be explained in terms of bias and variance? Now turn pruning back on in J48 and run the same experiment. Have the results changed? Explain why or why not, making a connection again to bias and variance.

2. (Gaussian Mixture Models – 10pts) What would the results be of modeling each of these data sets as a mixture of two Gaussians? Briefly justify your answer. If you do not believe it would have a good fit to the data, state one way we could modify our approach to improve the fit.



3. (Image Segmentation Using K-Means – 40pts) In this problem, you will apply K-Means to image segmentation. You are free to use whatever programming language you like for this problem. However, all of the instructions will be based on Matlab. If you choose to use another language, then you are responsible for figuring out the corresponding commands.

   Choose a nice natural image, such as a farmhouse against a blue sky, or a city scene. First load the image into Matlab using `imread` and convert it into floating point numbers (using `double`). The image is represented as a 3-D matrix of size image-width x image-height x 3. For each location in the image (x_i, y_i), the matrix contains 3 values for the red, green, and blue components of the pixels. We will use these pixel values for clustering. In addition to the pixel values $(r_i g_i b_i)$ for pixel i, we will also use the coordinates (x_i, y_i) as features. In particular, each data point f_i (i.e. pixel) should be a feature vector

$$f_i = \begin{bmatrix} r_i & g_i & b_i & x_i & y_i \end{bmatrix}$$

   To improve results, you should also standardize the values of each feature by computing the mean μ_c and standard deviation σ_c of each feature $c \in \{r, g, b, x, y\}$ over all pixels and setting $c'_i = \frac{\mu_c - c_i}{\sigma_c}$ for each entry c'_i in the feature vector. This ensures that each feature in the representation has a mean of 0 and a standard deviation of 1.

- Implement your own version of K-Means and use it to cluster the data (i.e. the features for each pixel) into 24 clusters (note that you can change this number once you're finished to try and get better results). If a cluster is empty during the procedure, assign a random data point to it. Use random initializations for the cluster centers, and iterate until the centroids converge.
- Use the cluster centers to generate the segmented image by replacing each data point with the closest center. Put the resulting data (only the color values, not the pixel position) back into the image data structure, i.e. the matrix representation where we started. Note that you also have to undo the feature standardization at this point (just solve the standardization equation for c_i given c_i'). Show the image using `imshow` and save the segmented image using `imwrite`.
- Include a page in your submission with your original image alongside your segmented image. Printing in grayscale is fine. The result of this process is called an over-segmented image. It is the first step to building such systems as this: `http://make3d.cs.cornell.edu/`. Later steps would piece these segments together into objects.
- Also print out your code and include it with your submission.

4. (K-Means – 10 pts) When using the K-Means clustering algorithm, we seek to minimize the variance of the solution. In general, what happens to the variance of a partition as you increase the value of K (the number of clusters) and why?

5. (Latent Variables and EM – 15 pts) [Adapted from Bishop Question 9.3] Consider a Gaussian mixture model in which the marginal distribution $p(z)$ for the latent variable is given by Equation 7.7, and the conditional distribution $p(x|z)$ for the observed variable is given by Equation 7.8. Show that the marginal distribution $p(x)$, obtained by summing $p(z)p(x|z)$ over all possible values of z, is a Gaussian mixture of the form
$$p(x) = \sum_{i=1}^{k} \pi_i \mathcal{N}(x \mid \mu_i, \Sigma_i) \ .$$

6. (Extra Credit Option A – 10 pts) [You may receive credit for only ONE of the extra credit problems.] With high-dimensional data we cannot perform visual checks, and problems can go unnoticed if we assume nice round, filled clusters. Describe in words a clustering algorithm which works even for weirdly-shaped clusters with unknown mixing ratios. You can assume that the clusters do not overlap, and that you have a LOT of unlabeled training data. Discuss the weaknesses of your algorithm. Do not work out the details for this problem; just convince me that you know the basic idea and understand its limitations. Please keep your answers *brief*. (Hint : Nothing weve covered so far in class will help you here; think about graphs....)

7. (Extra Credit Option B – 15 pts) [You may receive credit for only ONE of the extra credit problems.] Let $\mathbf{W}$ be a square $n \times n$ matrix. Is $d_{\mathbf{W}}(x_i, x_j) = ((x_i - x_j)^\intercal \mathbf{W}(x_i - x_j))^{\frac{1}{2}}$ still a metric (that is, does it satisfy the four properties of non-negativity, reflexivity, symmetry, and the triangle inequality)? Under what conditions on $\mathbf{W}$?