# Homework 5 Questions

Instructions

- 3 questions.

- Write code where appropriate.

- Feel free to include images or equations.

- Please make this document anonymous.

- **Please use only the space provided and keep the page breaks.** Please do not make new pages, nor remove pages. The document is a template to help grading. If you need extra space, please use and refer to new pages at the end of the document.

Questions

Q1: Given a linear classifier, how might we handle data that are not linearly separable? How does the *kernel trick* help in these cases? (See course slides in supervised learning, plus your own research.)

A1: The kernel trick helps us to separate data that are not linearly separable. It uses a mapping function called the kernel to transform the given space to a space of usually very high dimension. The idea is that if the data is not linearly separable in our *n* dimensional space, it may be separable in a higher dimension. An example of kernel transformation is shown in the figure below.
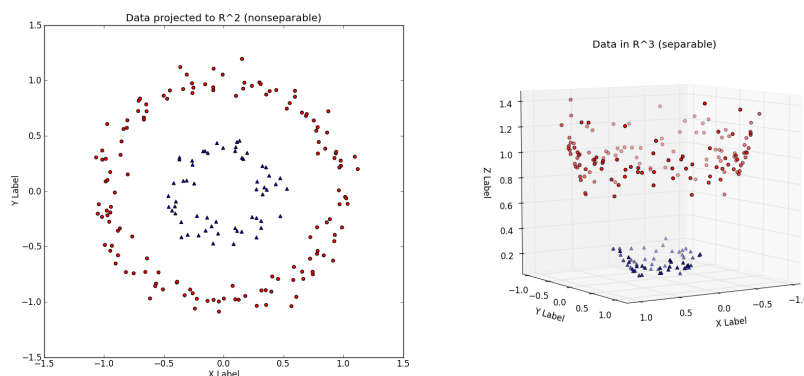
Q2:   In machine learning, what are bias and variance? When we evaluate a classifier, what are overfitting and underfitting, and how do these relate to bias and variance?

A2:   The bias error occurs from erroneous assumption in the learning algorithm. It is the difference between the prediction and the true value. Underfitting occurs when the model cannot adequately capture the underlying structure of the data. In underfitting, the bias error is high and the algorithm misses the relevant relations between features and target outputs. The variance is the sensitivity to small fluctuations in the training data. Overfitting occurs when the model matches the data too closely that it is likely to fail in additional data. When variance is high, the model matches the random noise in the training data as well, also known as overfitting.

Q3: The way that the bag of words representation handles the spatial layout of visual information can be both an advantage and a disadvantage. Describe an example scenario for each of these cases, plus describe a modification or additional algorithm which can overcome the disadvantage.

How might we evaluate whether bag of words is a good model?

A3: The bag of words method ignores the spatial relationships among the patches, but it can be important in an image representation. Such limitation can be overcome by including the spatial information into the algorithm. For example, the image can be divided into increasingly smaller sub-regions and compute local histograms in each sub-region as the pyramid matching approach. Also, the augmentation of SIFT by their spatial coordinates can introduce spatial information to the bag of words model. Although it does not consider the spatial relationship between patches, it is an intuitive way of classifying images based on the frequency of each visual words if they have unique words in each image. The bag of words model is commonly evaluated using the confusion matrix.