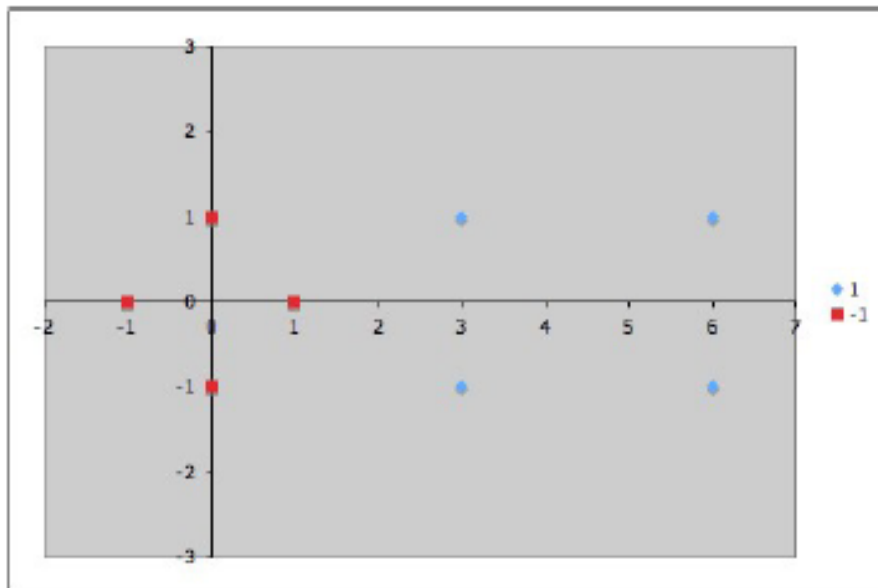


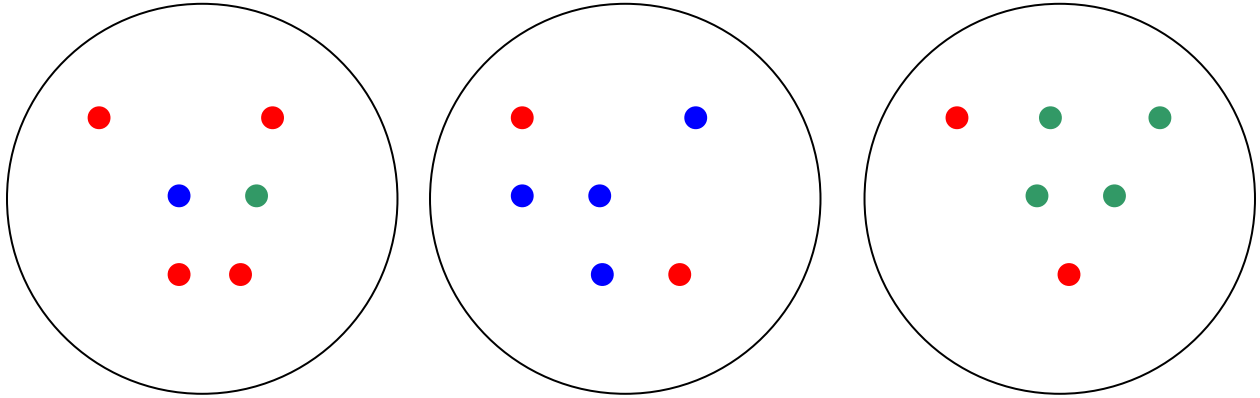
CS 599 - Information Retrieval, Homework 5

1. **(10 points)** For an SVM, if we remove one of the support vectors from the training set, does the size of the maximum margin decrease, stay the same, or increase for that dataset? Why? Also justify your answer by providing a simple dataset (no more than 2-D) in which you identify the support vectors, draw the location of the maximum margin hyperplane, remove one of the support vectors, and draw the location of the resulting maximum margin hyperplane.

2. **(30 points)** Apply linear SVM without soft margin to the following problem.
 - a. Please mark out the **support vectors**, the decision boundary $\mathbf{w}^T \mathbf{x} + \mathbf{b} = 0$ and $\mathbf{w}^T \mathbf{x} + \mathbf{b} = 1$ and $\mathbf{w}^T \mathbf{x} + \mathbf{b} = -1$. Note that you don't need to solve any optimization problem for this, just eyeball the solution (hyperplane) by looking at the following figure.
 - b. Solve for \mathbf{w} and \mathbf{b} based on the support vectors.
 - c. Calculate the margin of the support vector using the \mathbf{w} .



3. **(40 points)** Consider the following 3 clusters.
 - a. Calculate the Purity
 - b. Create the contingency table and using the contingency table,
 - i. Calculate the Rand index
 - ii. Calculate the Balanced F measure



4. **(20 points)** Describe your thoughts about what you think it means to work as a software engineer (or computer scientist). You may therefore – if you like – be very personal and describe your own plans and fears for your future career, criticism of your education, skills you need to develop further, and soon. This homework question is intended to encourage you to reflect about yourself and your future career, and will therefore be graded generously!