Department of Computer Science, University of Houston
COSC 4335 - Exercise set 1
Spring 2012, Due Tue. 21, 2012 at 11.30am.
Staple all sheets together. No unstapled homeworks will be accepted.

Remember the academic honesty policy for the course. All work turned in **must** be your own. Copying solutions from anyone else or working on problems together is disallowed. Searching/downloading from internet or other text books for solutions is disallowed unless explicitly allowed for a question.

1. Section 1.7. Exercise 2.

2. Section 2.6. Exercise 2 parts (a), (d), (g), (j).

3. Section 2.6. Exercise 3. Answer to (a): Boss is right.

4. Section 2.6. Exercise 6.

5. Section 2.6. Exercise 14. Answer: Euclidean after transformation.

6. Section 2.6. Exercise 20 part (b).

7. Which of the similarity measures discussed in Section 2.4 are appropriate for text documents. Justify your answers. Hint: We have already discussed one in class: the cosine function.

8. For the Iris data set, perform the indicated operations each time starting with the original values and **compare** the results for the new data with that for the original data. First, multiply column C values by 5 and column D values by 5 and then recalculate r, the pearson correlation coefficient. Second, square columns C and D and then recalculate r. Finally, square only the C column values and then recalculate r for the new column C and original column D. Explain your observations. Hint: consider the formula for r.