

Homework 1, 290D, Fall 07

Due October 15, 5 pm

- (5 points) Consider the Hamming distance between two strings defined as the number of mismatched positions in the two strings. Show that this distance is a metric. You may assume that the strings are of equal length.
- (10 points) Consider the edit distance between two strings defined as the minimum number of inserts, deletes, and replacement operations needed to transform one string into another. Show that this distance is a metric.
- (15 points) Consider the dynamic time warp distance between two sequences of values x and y that is defined as follows.
$$d(x,y) = 0 \quad \text{if } x \text{ and } y \text{ are both null}$$
$$d(x,y) = \infty \quad \text{if one of } x \text{ and } y \text{ is null}$$
otherwise, let $x = b.x'$ and $y = c.y'$.
$$d(x,y) = |b-c| + \min\{d(x',y), d(x,y'), d(x',y')\}$$
What is the time complexity of computing d ? Is this a metric?
- (5 points) Extend the vectors $(3,2,5)$, $(1,2,3)$ into a basis for 3-d Euclidean space. Show that your vectors span the space.
- (10 points) Find a basis for the following subspaces of \mathbb{R}^4 .
 - The vectors for which $x_1 = 2x_4$.
 - The vectors for which $x_1+x_2+x_3 = 0$ and $x_3+x_4 = 0$.
- (5 points) Can the intersection of two subspaces of a vector space be empty? Explain.
- (20 points) Suppose vector space V has a dimension of k . Prove that
 - any k independent vectors in V form a basis
 - any k vectors that span V form a basis.
- (30 points) This problem compares the cost of accessing data in memory with that of performing random accesses and sequential accesses on disk. Generate a binary file of 100 million doubles (or download a sample one approx 800MB from <http://amazon.cs.ucsb.edu/CS290D/>). Implement three separate procedures and time their execution as follows:
 - Start the timing. Open the file at position 1000 read 1 million consecutive doubles
 - Compute 1 million random numbers between 0 and 10 million . Start the timing. Read the doubles at those positions
 - Load 1 million doubles in a double[] . Start the timing. Iterate through the array.

Report the execution times in the three cases.