CS 590D: Data Mining

Spring 1998

Problem Set 3: Associations and Correlations

Due Date: Apr 21, 1997 in class

No late submissions will be accepted

1. Perform a time-series analysis to predict the next 5 entries of the following series (use variable step sizes and windows). The actual entries are 1.681061 0.628649 -0.322779 0.366148 -1.078785. What is the percentage absolute error for your best prediction and the real numbers?

   0.000000 1.750768 0.152495 -0.138295 0.232556 -1.502945 -0.815988 1.647594 0.701455 -0.338869 0.368924 -1.008842 -1.442151 1.182725 1.261513 -0.337744 0.263523 -0.432315 -1.742766 0.446246 1.658058 -0.079866 0.008851 0.055568 -1.673833 -0.394727 1.749186 0.397587 -0.250645 0.329239 -1.292842 -1.143218 1.471453 0.973361 -0.368845 0.345708 -0.737955 -1.628684 0.862476 1.477774 -0.248775 0.154606 -0.183331 -1.755233 0.053100 1.744900 0.122322 -0.121679 0.215333 -1.527135 -0.7687 40 1.665056 0.665005 -0.331217 0.368029 -1.043998 -1.411547 1.221145 1.229534 -0.345214 0.275801 -0.467405 -1.734868 0.497347 1.641064 -0.103277 0.026532 0.030405 -1.688361 -0.342837 1.754130 0.363260 -0.237198 0.319697 -1.323480 -1.102658 1.499428 0.937600 -0.367806 0.351711 -0.774463 -1.608338 0.908162 1.451656 -0.263983 0.170574 -0.214799 -1.758280 0.106151 1.737645 0.092844 -0.104794 0.197118 -1.550306 -0.720776

   (you can get the series on your class web site online.)

2. Write a program to compute all implications with a confidence of 0.7 and support of 0.1 in the following transaction set. {{a, b, c, d, f}, {a, f, g}, {a, c, g, h, k}, {b, d, e, k}, {e, l, m}, {l, n, p}, {a, p, q}, {b, q, r}, {c, q, r}, {p, r, t}, {d, p, r, s, t, u}, {s, t, u}, {b, s, t, u}, {u, v, z}, {u, v, w}, {w, z}}

3. Write a program to compute all correlated sets in a chi-squared sense on the same dataset. Use a chi-squared significance of 3.5, support of 0.2, support fraction p of 0.3. Use the algorithm discussed in the class and in the paper "Beyond Market Baskets: Generalizing Association Rules to Correlations., S. Brin, R. Motwani and C. Silverstein, 1997 ACM SIGMOD Conference on Management of Data".