Data Mining Angabe aus 2017 18

Data Mining und Knowledge Discovery (IN2030) (Technische Universität München)

Data Mining Exam 2017/18

Problem 1

There are several methods for data analysis. Given X, a set of numerical data, when do we use each of these methods:

a) Principal component analysis, if X…

b) Median filter, if X…

c) Fourier analysis, if X...

d) Fuzzy relational c-means, if X...

e) ID3, if X...

f) Cosine similarity, if X…

g) Edit distance, if X...

Problem 2

Define all values of x such that r(x,(0,0))=1 where r is:

a) Euclidean distance

b) City block distance

c) Inner product defined by the norm inducing matrix A = $\begin{bmatrix} 0.1 & 0 \\ 0 & 1 \end{bmatrix}$

d) Inner product defined by the norm inducing matrix A = $\begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}$

e) Hamming distance

f) Edit distance

Problem 3

Given two classes X_1 = {1,2,3} and X_2 = {4+$\sqrt{2}$,4}, where cov(X_1) = cov(X_2) =1 , define the classification border when using:

a) Linear discriminant analysis

b) SVM with no kernelization

c) 3-nearest neighbours

Problem 4

For X={(2,0),(0,2)} you have 2 clusters (c=2)

a) If V is initialized to V={(0,0),(1,0)}, which partition matrices and which V updates will this initialization yield until convergence?

b) Now if V_1 = (0,0) and V_2 = (-1,0), which partition matrices and which V updates will this initialization yield until convergence?

c) State the values for V_2 such that the second cluster is empty