

Algoritmos e Estruturas de Dados 2

Primeiro Trabalho Prático (Segunda Versão)

Entrega Sugerida: 13/08/2023

1º Semestre 2023 - DC-UFSCar

1 Introdução

No primeiro Trabalho Prático (TP01) será solicitada a entrega de um programa que solucione o problema apresentado na próxima seção. O arquivo entregue deve seguir os itens abaixo:

- O TP01 deverá ser feito individualmente e plágio não será tolerado;
- O TP01 deve ser feito em linguagem C, com um cabeçalho com as informações do estudante (nome, curso, RA);
- Compile o seu TP01 usando o compilador GCC com flags -Wall -pedantic -O2 -Wno-unused-result.

2 Detecção de Outliers

A detecção de outliers é uma atividade muito comum em diversas áreas, como a Estatística e a Mineração de Dados. Um outlier é um objeto cujos valores de seus atributos se difere muito da maioria dos outros objetos observados. Calma, jovem. Já vamos explicar melhor.

Um objeto pode ser qualquer tipo de entidade, descrita por atributos. Atributos são características que descrevem esse objeto. Por exemplo, uma pessoa pode ser considerada um objeto. Seus atributos seriam, entre outros, sua altura, seu peso e sua idade. Uma pessoa com 2,35 metros de altura, por exemplo, pode ser considerada um outlier.

A definição de outlier pode variar de uma área para outra, ou até mesmo de um autor para outro, na literatura. Neste trabalho, utilizaremos uma definição que é muito comum na literatura de Mineração de Dados. Um outlier é aquele objeto cuja distância até o seu k -ésimo vizinho mais próximo é estritamente maior que um limiar L . Ou seja, se a distância para o k -ésimo objeto mais próximo de um objeto A ultrapassar um valor previamente definido L , então A pode ser considerado um outlier. Lembre que o primeiro vizinho mais próximo é o 0-ésimo.

A distância entre dois objetos pode ser calculada de várias maneiras. Neste trabalho, considere que essas distâncias já foram calculadas e são parte da entrada.

Entrada

Cada caso de teste é iniciado por uma linha contendo um inteiro N ($2 \leq N \leq 10^9$), um inteiro S ($1 \leq S \leq 1000$), um inteiro k ($0 \leq k < N$) e um real positivo L . O número total de objetos no conjunto que estamos considerando é indicado por N . Porém, em função de restrições no tamanho das entradas, só vamos procurar outliers entre os primeiros S objetos desse conjunto. Assim, temos S linhas após a primeira, e cada uma contém N números. Vamos associar um índice i , entre 0 e $S - 1$ a cada linha após a primeira. Especificamente, a i -ésima linha contém a distância do i -ésimo objeto para todos os demais. Assim, associemos um índice j , entre 0 e $N - 1$, a cada valor de uma determinada linha. Dessa forma, $D_{i,j}$ é o j -ésimo valor da linha i e indica a distância de i até j . Na entrada, é garantido que $D_{i,i} = 0$, ou seja, a distância de um objeto a si mesmo é zero.

Saída

A saída de um caso de teste contém apenas uma linha (terminada em '\n') contendo um inteiro, que indica o número de outliers encontrados.

Exemplos de entradas e saídas correspondentes

Entrada	Saída
5 5 0 0.7 0.0 1.0 1.5 3.0 2.0 1.0 0.0 2.0 5.0 1.5 2.2 1.7 0.0 2.3 0.7 1.8 0.3 0.1 0.0 2.4 1.5 1.6 2.3 0.7 0.0	2

Entrada	Saída
5 2 1 1.3 0.0 1.0 1.5 3.0 2.0 1.0 0.0 2.0 5.0 1.5	2