

Algoritmos e Estruturas de Dados 2

Segundo Trabalho Prático (Segunda Versão)

Entrega Sugerida: 19/10/2019

2º Semestre 2019 - DC-UFSCar

1 Introdução

No segundo Trabalho Prático (TP02) será solicitada a entrega de um programa que solucione o problema apresentado na próxima seção. O arquivo entregue deve seguir os itens abaixo:

- O TP02 deverá ser feito individualmente e plágio não será tolerado;
- O TP02 deve ser entregue no run codes (<https://run.codes>) em um arquivo contendo código em uma das seguintes linguagens (C, C++, Java) e com um cabeçalho com as informações do estudante (nome, curso, RA);
- Cada estudante deve se cadastrar no run codes (<https://run.codes>) informando Nome Completo, escolhendo “UFSCar - Universidade Federal de São Carlos” no campo Universidade e colocando seu RA no campo Núm. Matrícula. Depois de cadastrado, basta logar no run codes e se matricular na disciplina “1001490 - Algoritmos e Estruturas de Dados 2” usando o Código de Matrícula 9T4K.
- Compile o seu TP usando o compilador GCC com flags -Wall -pedantic, pois warnings podem impedir o código de funcionar no run codes ainda que funcionem no seu computador;

2 Detecção de Outliers

A detecção de outlier 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 não seja menor 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.

Entrada

Cada caso de teste está armazenado em um arquivo com extensão “.in”. A única linha a ser lida da entrada padrão contém o nome do arquivo contendo a instância. Este arquivo é iniciado por uma linha contendo um inteiro N ($2 \leq N \leq 10^9$), um inteiro k ($0 \leq k < n$) e um real positivo L . As próximas N linhas contém N números reais cada, indicando as distâncias $D_{i,j}$ entre os objetos i e j . Especificamente, a i -ésima linha representa a distância do i -ésimo objeto para todos os demais. Na entrada, é garantido que $D_{i,i} = 0$, ou seja, a distância de um objeto a si mesmo é zero. O arquivo “modeloFILE.c” tem um exemplo de como fazer essa leitura.

Saída

A saída de um caso de teste contém apenas uma linha contendo um inteiro, que representa o número de outliers naquele conjunto de objetos.

Exemplo de entrada

```
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
```

Saída esperada para esse exemplo

2