

# P-Center em Julia

February 12, 2016

## 1 Trabalho de Implementação

### 1.1 INF2912 - Otimização Combinatória

1.1.1 Prof. Marcus Vinicius Soledade Poggi de Aragão

1.1.2 2015-2

1.1.3 Ciro Cavani

BigData / Globo.com Algoritmos de clusterização.

### 1.2 Conteúdo

Esse notebook tem o desenvolvimento e avaliação do algoritmo aproximado do P-Center (algoritmo Farthest-first traversal).

A avaliação do algoritmo é baseada em um mapeamento entre a maioria dos itens que foram atribuídos a um determinado cluster e o correspondente os valores verdadeiros gerados nesse cluster.

O P-Center teve resultados muito bons.

### 1.3 Dataset

```
In [1]: include("../src/clustering.jl")
import Inf2912Clustering
const Clustering = Inf2912Clustering
dataset = Clustering.load_dataset("small")
Clustering.summary(dataset)
sleep(0.2)
```

WARNING: type Dataset not present in workspace; reconstructing

```
LoadError: MethodError: ‘summary’ has no method matching summary(::JLD.##Dataset#8091)
you may have intended to import Base.summary
while loading In[1], in expression starting on line 5
```

#### 1.3.1 P-Center - Problema de Localização de Centróides

Consiste em resolver o P-Center determinar os objetos representantes de cada grupo e classificar cada objeto como sendo do grupo com representante mais próximo

[https://en.wikipedia.org/wiki/Metric\\_k-center](https://en.wikipedia.org/wiki/Metric_k-center)

[https://en.wikipedia.org/wiki/Farthest-first\\_traversal](https://en.wikipedia.org/wiki/Farthest-first_traversal)





```

        assignments = pcenter(dataset, k)
        centermap = Clustering.mapping(dataset, assignments, k)
        map(c -> centermap[c], assignments)
    end

    let
        n = 100
        k = 3
        c = 16
        c_y = 3

        tiny = Clustering.Dataset(size=n, groups=k, features=c, slot=c_y)

        prediction = pcenter_approx(tiny, k)
        Clustering.evaluation_summary(tiny, prediction; verbose=true)
    end

```

Matriz de Confusão:

```

[23 8 11
 14 18 3
 2 7 14]

```

Tamanho: 100  
 Acertos: 55  
 Erros: 45  
 Accuracy: 55.0%

Cluster 1

Tamanho: 42  
 Accuracy: 65.0%  
 Precision: 58.97%  
 Recall: 54.76%  
 F-score: 0.57

Acerto positivo: 23 (54.76%)  
 Acerto negativo: 42 (72.41%)  
 Falso negativo: 19 (42.22%)  
 Falso positivo: 16 (35.56%)

Cluster 2

Tamanho: 35  
 Accuracy: 68.0%  
 Precision: 54.55%  
 Recall: 51.43%  
 F-score: 0.53

Acerto positivo: 18 (51.43%)  
 Acerto negativo: 50 (76.92%)  
 Falso negativo: 17 (37.78%)  
 Falso positivo: 15 (33.33%)

### Cluster 3

Tamanho: 23

Accuracy: 77.0%

Precision: 50.0%

Recall: 60.87%

F-score: 0.55

Acerto positivo: 14 (60.87%)

Acerto negativo: 63 (81.82%)

Falso negativo: 9 (20.0%)

Falso positivo: 14 (31.11%)