

Análise

February 14, 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

TBD

1.3 Dataset

```
In [1]: include("../src/clustering.jl")
import Inf2912Clustering
const Clustering = Inf2912Clustering
```

WARNING: redefining constant srcdir

WARNING: redefining constant default_datasetdir

```
Out[1]: Inf2912Clustering
```

```
In [2]: function validate_prediction(dataset, prediction)
        size = length(prediction)
        if size != dataset.size
            error("Mismatch size: expected $(dataset.size), actual $(size)")
        end
        clusters = length(unique(prediction))
        if clusters != dataset.clusters
            error("Mismatch clusters: expected $(dataset.clusters), actual $(clusters)")
        end
    end
```

```
Out[2]: validate_prediction (generic function with 1 method)
```

```
In [3]: import Clustering: mapping, Evaluation

        function run(predictor; iters=10, approx=true)
            for i=1:iters
                println("Iteration ", i, "\n")
                dataset = Clustering.dataset_tiny()
```

```

        k = dataset.clusters
        print("Time: ")
        @time prediction = predictor(dataset, k)
        try
            if approx
                centermap = mapping(dataset, prediction, k)
                prediction = map(c -> centermap[c], prediction)
            end
            validate_prediction(dataset, prediction)
            r = Evaluation(dataset, prediction)
            print(r.sample)
        catch e
            println("Error: ", e.msg)
        end
        println()
    end
end

```

Out[3]: run (generic function with 1 method)

In [4]: run((dataset, k) -> Clustering.random_clustering(dataset); approx=false)
sleep(0.2)

Iteration 1

Time: 0.057064 seconds (29.43 k allocations: 1.493 MB)
Size: 100
Correct: 33
Mistakes: 67
Accuracy: 33.0%

Iteration 2

Time: 0.000059 seconds (55 allocations: 6.344 KB)
Size: 100
Correct: 33
Mistakes: 67
Accuracy: 33.0%

Iteration 3

Time: 0.000050 seconds (55 allocations: 6.297 KB)
Size: 100
Correct: 28
Mistakes: 72
Accuracy: 28.0%

Iteration 4

Time: 0.000052 seconds (56 allocations: 6.891 KB)
Size: 100
Correct: 30
Mistakes: 70
Accuracy: 30.0%

Iteration 5

Time: 0.000040 seconds (55 allocations: 6.344 KB)
Size: 100
Correct: 34
Mistakes: 66
Accuracy: 34.0%

Iteration 6

Time: 0.000036 seconds (55 allocations: 6.344 KB)
Size: 100
Correct: 34
Mistakes: 66
Accuracy: 34.0%

Iteration 7

Time: 0.000037 seconds (56 allocations: 6.891 KB)
Size: 100
Correct: 27
Mistakes: 73
Accuracy: 27.0%

Iteration 8

Time: 0.000036 seconds (56 allocations: 6.891 KB)
Size: 100
Correct: 31
Mistakes: 69
Accuracy: 31.0%

Iteration 9

Time: 0.000032 seconds (56 allocations: 6.891 KB)
Size: 100
Correct: 35
Mistakes: 65
Accuracy: 35.0%

Iteration 10

Time: 0.000032 seconds (56 allocations: 6.938 KB)
Size: 100
Correct: 43
Mistakes: 57
Accuracy: 43.0%

```
In [5]: run(Clustering.kmeans)
        sleep(0.2)
```

Iteration 1

Time: 0.516949 seconds (591.60 k allocations: 26.575 MB, 1.31% gc time)
Size: 100

Correct: 92
Mistakes: 8
Accuracy: 92.0%

Iteration 2

Time: 0.001075 seconds (5.67 k allocations: 427.063 KB)
Error: Center already mapped: 1 -> 1, now 2?

Iteration 3

Time: 0.001845 seconds (10.01 k allocations: 750.266 KB)
Size: 100
Correct: 84
Mistakes: 16
Accuracy: 84.0%

Iteration 4

Time: 0.003179 seconds (14.36 k allocations: 1.051 MB)
Size: 100
Correct: 82
Mistakes: 18
Accuracy: 82.0%

Iteration 5

Time: 0.001269 seconds (7.11 k allocations: 535.219 KB)
Size: 100
Correct: 95
Mistakes: 5
Accuracy: 95.0%

Iteration 6

Time: 0.001229 seconds (7.12 k allocations: 536.313 KB)
Size: 100
Correct: 83
Mistakes: 17
Accuracy: 83.0%

Iteration 7

Time: 0.002648 seconds (15.80 k allocations: 1.154 MB)
Size: 100
Correct: 87
Mistakes: 13
Accuracy: 87.0%

Iteration 8

Time: 0.002767 seconds (11.46 k allocations: 858.141 KB)
Size: 100
Correct: 87

Mistakes: 13
Accuracy: 87.0%

Iteration 9

Time: 0.001841 seconds (10.00 k allocations: 748.125 KB)
Error: Center already mapped: 1 -> 1, now 3?

Iteration 10

Time: 0.001503 seconds (8.56 k allocations: 644.016 KB)
Size: 100
Correct: 93
Mistakes: 7
Accuracy: 93.0%

```
In [6]: run(Clustering.pcenter)
        sleep(0.2)
```

Iteration 1

Time: 0.269964 seconds (293.23 k allocations: 12.086 MB, 2.57% gc time)
Size: 100
Correct: 60
Mistakes: 40
Accuracy: 60.0%

Iteration 2

Time: 0.000396 seconds (3.14 k allocations: 188.938 KB)
Error: Center already mapped: 2 -> 1, now 2?

Iteration 3

Time: 0.000382 seconds (3.14 k allocations: 188.938 KB)
Size: 100
Correct: 64
Mistakes: 36
Accuracy: 64.0%

Iteration 4

Time: 0.000381 seconds (3.14 k allocations: 188.938 KB)
Error: Center already mapped: 1 -> 1, now 3?

Iteration 5

Time: 0.000422 seconds (3.14 k allocations: 188.938 KB)
Size: 100
Correct: 57
Mistakes: 43
Accuracy: 57.0%

Iteration 6

Time: 0.000380 seconds (3.14 k allocations: 188.938 KB)
Size: 100
Correct: 68
Mistakes: 32
Accuracy: 68.0%

Iteration 7

Time: 0.000375 seconds (3.14 k allocations: 188.938 KB)
Size: 100
Correct: 61
Mistakes: 39
Accuracy: 61.0%

Iteration 8

Time: 0.000376 seconds (3.14 k allocations: 188.938 KB)
Size: 100
Correct: 63
Mistakes: 37
Accuracy: 63.0%

Iteration 9

Time: 0.000378 seconds (3.14 k allocations: 188.938 KB)
Size: 100
Correct: 54
Mistakes: 46
Accuracy: 54.0%

Iteration 10

Time: 0.000375 seconds (3.14 k allocations: 188.938 KB)
Error: Center already mapped: 1 -> 1, now 2?

```
In [7]: run(Clustering.pmedian)
        sleep(0.2)
```

Iteration 1

Time: 4.379394 seconds (4.21 M allocations: 187.112 MB, 1.35% gc time)
Size: 100
Correct: 84
Mistakes: 16
Accuracy: 84.0%

Iteration 2

Time: 0.912891 seconds (112.72 k allocations: 10.313 MB, 1.00% gc time)
Size: 100
Correct: 81
Mistakes: 19
Accuracy: 81.0%

Iteration 3

Time: 0.846140 seconds (110.19 k allocations: 10.211 MB)
Size: 100
Correct: 86
Mistakes: 14
Accuracy: 86.0%

Iteration 4

Time: 0.974865 seconds (110.19 k allocations: 10.211 MB, 0.69% gc time)
Size: 100
Correct: 85
Mistakes: 15
Accuracy: 85.0%

Iteration 5

Time: 1.212883 seconds (110.19 k allocations: 10.211 MB)
Size: 100
Correct: 84
Mistakes: 16
Accuracy: 84.0%

Iteration 6

Time: 1.005441 seconds (110.19 k allocations: 10.211 MB, 0.53% gc time)
Size: 100
Correct: 82
Mistakes: 18
Accuracy: 82.0%

Iteration 7

Time: 7.897316 seconds (110.19 k allocations: 10.211 MB)
Size: 100
Correct: 84
Mistakes: 16
Accuracy: 84.0%

Iteration 8

Time: 1.081709 seconds (110.19 k allocations: 10.211 MB, 0.54% gc time)
Size: 100
Correct: 80
Mistakes: 20
Accuracy: 80.0%

Iteration 9

Time: 0.960425 seconds (110.19 k allocations: 10.211 MB)
Size: 100
Correct: 87
Mistakes: 13
Accuracy: 87.0%

Iteration 10

Time: 1.114462 seconds (110.19 k allocations: 10.211 MB, 0.46% gc time)

Size: 100

Correct: 78

Mistakes: 22

Accuracy: 78.0%