

## Atividade 1

# Machine Learning - Clustering

---

Natália Sales Mesquita 21341077

Professor Felipe Kopel

São Paulo, 08 Agosto de 2019

**1) Indique um exemplo de distância (dissimilaridade) não euclidiana e sua aplicação. Cite referência (.edu; artigo científico)**

Em estatística, a distância de Mahalanobis é uma medida de distância introduzida pelo matemático indiano Prasanta Chandra Mahalanobis em 1936.[1] É baseada nas correlações entre variáveis com as quais distintos padrões podem ser identificados e analisados. É uma estatística útil para determinar a similaridade entre uma amostra desconhecida e uma conhecida. Distingue-se da distância euclidiana já que tem em conta as correlações do conjunto de dados e é invariante à escala, ou seja, não depende da escala das medições.

```
[2] distancia = sqrt((x - y)' V^-1 (x - y))
```

Fontes:

[1] Multivariate Statistical Methods, J. Manly (ISBN 0-412-28620-3)

[2] <[https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.DistanceMetric.html#sklearn.neighbors.DistanceMetric.get\\_metric](https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.DistanceMetric.html#sklearn.neighbors.DistanceMetric.get_metric)> Acesso em Agosto/2019

**2) Calcule a tabela de distâncias (dissimilaridade) MINKOWSKI com n=1 n=3 nos pontos.**

	X1	X2	X3
P1	-3	0	6
P2	-1	-1	-1
P3	3	4	3
P4	10	0	0
P5	0	9	10

Código no Python

```
from scipy.spatial.distance import pdist

P1 = [-3, 0, 6]
P2 = [-1, -1, -1]
P3 = [ 3, 4, 3]
P4 = [10, 0, 0]
P5 = [ 0, 9, 10]

X = [P1, P2, P3, P4, P5]

Y = pdist(X, 'minkowski', p=1)
print(Y)
[10. 13. 19. 16. 13. 13. 22. 14. 15. 29.]
```

```
Y = pdist(X, 'minkowski', p=3)
print(Y)
[ 7.06069667  6.74599671 13.41278945  9.35990162  6.32470354 11.00550688
 13.26099716  7.57117428  7.91045989 13.97444317]
```