

UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
INSTITUTO METROPOLE DIGITAL
IMD1101 - Aprendizado de Máquina – 2024.2
Aula08 – Redução de Dimensionalidade

Visando exercitar os conceitos aprendidos na aula de Correlação de Pearson e Redução de Dimensionalidade, baixe o arquivo compactado que foi publicado na turma virtual (ver Figura 1), e que contém três bases que deverão ser pré-processadas:

Integrado de Gestão de Atividades Acadêmicas

rar Vínculo
semestre atual: 2024.2
IMD1101 - APRENDIZADO DE MÁQUINA - T01 (2024.2)

Aula07: Pré-processamento (redução de dimensionalidade 02) (10/10/2024 - 10/10/2024)

Aula08: Pré-processamento (prática de PCA e seleção de atributos) (11/10/2024 - 11/10/2024)

[Bases_Pratica_PCA.zip](#)

Figura 1. Arquivo compactado contendo três bases de dados em CSV.

De posse das bases, escreva um notebook em Python (Colab) que mostre a correlação de Pearson entre os atributos, de acordo com a Figura 2.

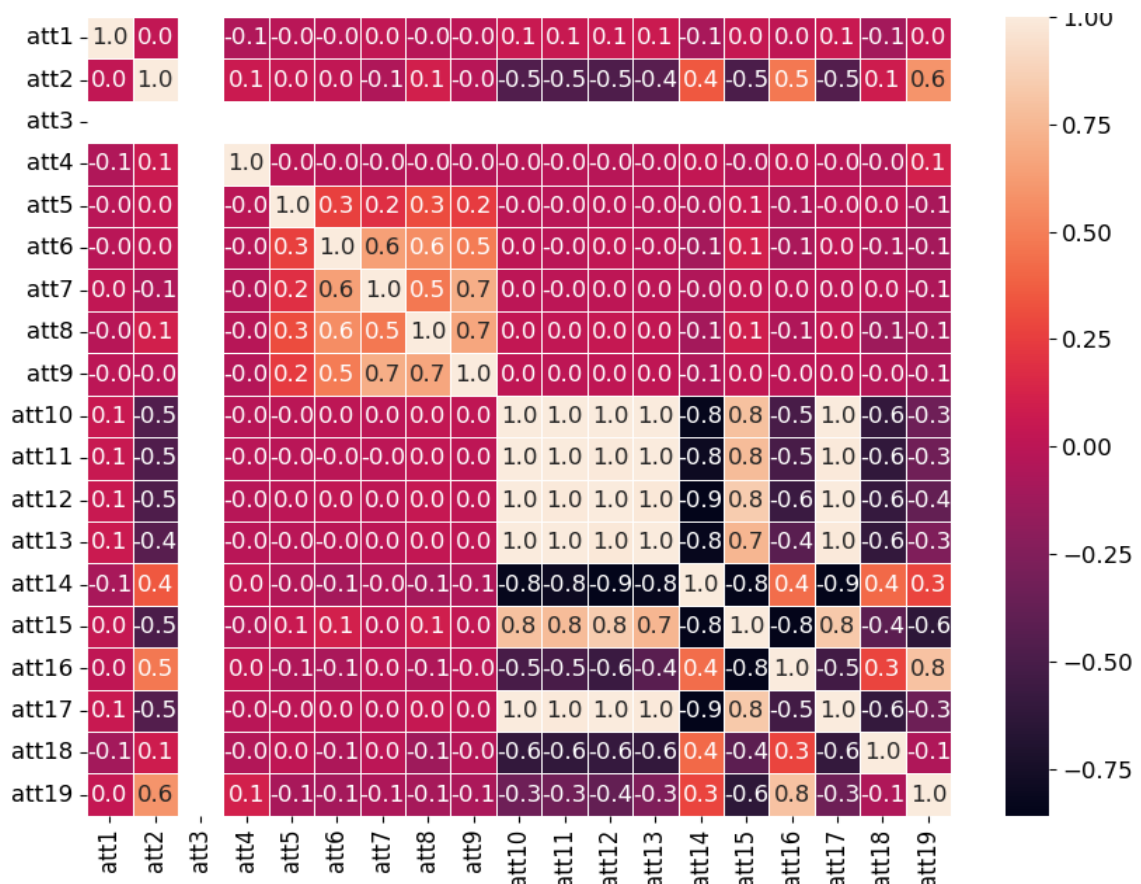


Figura 2. Visualização da Correlação de Pearson – Base Image.

Baseado no mapa de calor da Figura 2, escolha os dez atributos menos correlacionados de cada base.

A seguir, aplique dois métodos de PCA sobre as três bases, de tal forma que as bases resultantes apresentem apenas 10 componentes (primeiro método) ou que mantenham 90% da informação original (segundo método), de acordo com as Figuras 3 e 4.

```
[[ -69.10041529  18.16888289 -7.23297036 ... -1.28805147 -0.41879867
   0.14977705]
 [ -67.51374932  66.63786402 -8.77970528 ... -1.37520725 -0.18833928
   0.64981897]
 [ -77.87607376 -15.02576365 -6.54527403 ... -1.34925039 -0.46337563
   0.16850186]
 ...
 [ -78.92257926   9.96876337 -8.03673203 ... -1.88084963 -0.51297262
  -1.411826 ]
 [ -73.31367943  31.38483119 -7.57854351 ... -0.53949106  0.62148979
  -1.29221284]
 [ -72.83388219   5.80468646 -7.10009592 ... -0.57668728 -0.37262796
  -1.3054242 ]]
```

(2310, 10)

Figura 3. Visualização da base resultante com 10 componentes.

```
[[ -0.72266253  0.24910291 -0.10512707 -0.61986091]
 [ -0.70606894  0.91363273 -0.12760798 -0.54470738]
 [ -0.81443969 -0.20600945 -0.0951318  -0.29450686]
 ...
 [ -0.8253842   0.13667588 -0.11680929  0.12823458]
 [ -0.76672548  0.43029904 -0.11014978  0.05531865]
 [ -0.76170769  0.07958466 -0.10319582  0.2128716 ]]
```

(2310, 4)

Figura 4. Visualização da base resultante com 90% das informações.

Depois de aplicados os métodos de correlação e PCA, salve as bases (nove), três de correlação e seis de PCA, para serem utilizadas em atividades posteriores. Por último, submeta, via SIGAA, o seu Jupyter Notebook que executa todos os passos descritos acima. Sua submissão valerá a presença referente a essa aula assíncrona.

Bom trabalho!