

Breast Cancer Wisconsin (Diagnostic)

Luís Gustavo
Aramys Almeida Matos

Inteligência Computacional

23 de outubro de 2016

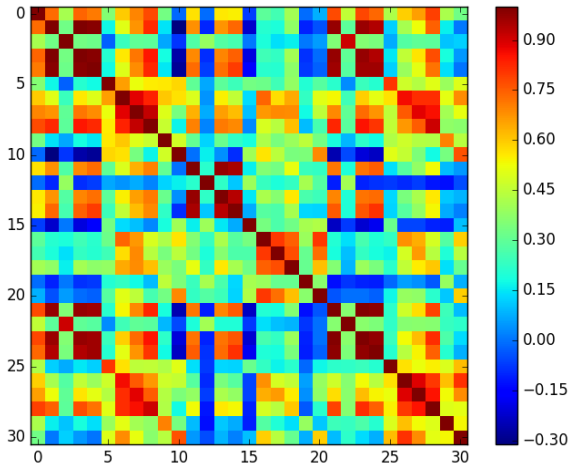
- Exames para câncer de mama.
- Características do núcleo de células computadas a partir de imagens digitalizadas
- 30 variáveis de entrada
- 1 variável de saída
- Raio, textura, perímetro, área, suavidade, compactidade, concavidade, pontos côncavos, simetria, dimensão fractal

- Raio
- textura
- perímetro
- área
- suavidade
- compacidade
- concavidade
- pontos côncavos
- simetria
- dimensão fractal

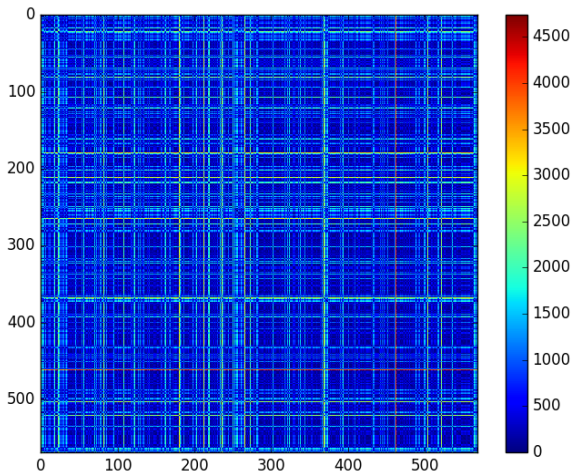
Para cada um desses dados são calculadas as métricas média, desvio padrão e pior valor, totalizando 30 variáveis.

- Problema de classificação
- Identificar se o câncer é benigno ou maligno.
- Aplicar e avaliar os modelos de classificação:
 - Classificador Bayesiano Simples
 - Classificador Bayesiano Quadrático
 - Mínimos Quadrados
 - Regressão Logística

Matriz de Correlação

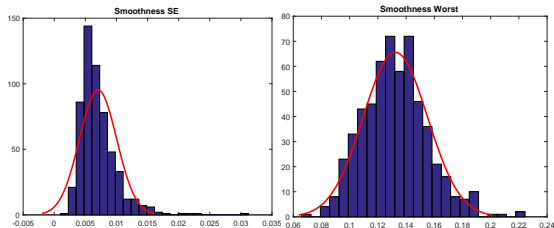
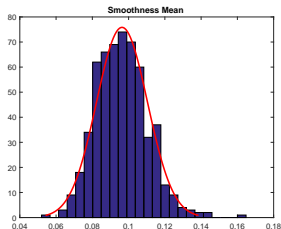


Distância Euclidiana



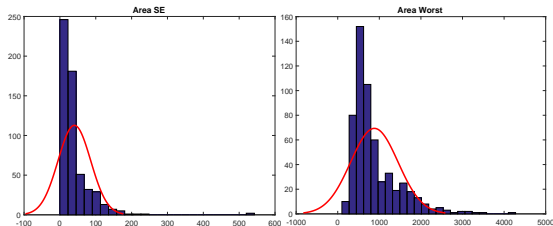
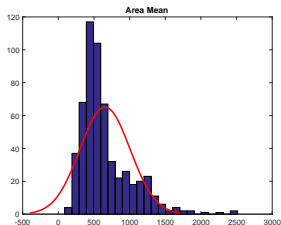
Histogramas

Smoothness



Histogramas

Area



Python

- SciPy
 - *NumPy*
 - *matplotlib*
 - *pandas*: Python Data Analysis Library
- *scikit-learn*: Machine Learning in Python

Matlab

- Statistics and Machine Learning Toolbox