



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

INTELIGÊNCIA COMPUTACIONAL

Relatório Preliminar

Nomes:

Aramys Almeida Matos

Luís Gustavo Oliveira Silva

Professor:

Alexandre Evsukoff

1 Introdução

Este trabalho será desenvolvido sobre o *dataset Breast Cancer Wisconsin (Diagnostic)*. Este *dataset* possui como variáveis características computadas a partir de imagens digitalizadas de exames de mama por punção aspirativa por agulha fina. Ele descreve características do núcleo das células presentes na imagem.

2 Caracterização

O dataset possui os seguintes dados.

- ID do paciente.
- Diagnóstico (Maligno ou benigno codificados como -1 e 1 respectivamente)

Atributos das células:

1. Raio (média das distâncias do centro para pontos no perímetro)
2. Textura
3. Perímetro
4. Área
5. Suavidade
6. Compacidade
7. Concavidade
8. Pontos côncavos
9. Simetria
10. Dimensão Fractal

Para esses 10 atributos é calculado média, erro padrão e o pior valor, resultando em 30 variáveis de entrada.

O dataset é bem balanceado, apresentando 357 amostras benignas, 212 amostras malignas.

3 Formulação do Problema

Consiste em um problema de classificação. Deseja-se desenvolver um classificador capaz de predizer a classe do registro (maligno ou benigno).

Para solução deste problema serão aplicados os modelos de classificação a seguir e avaliados os resultados.

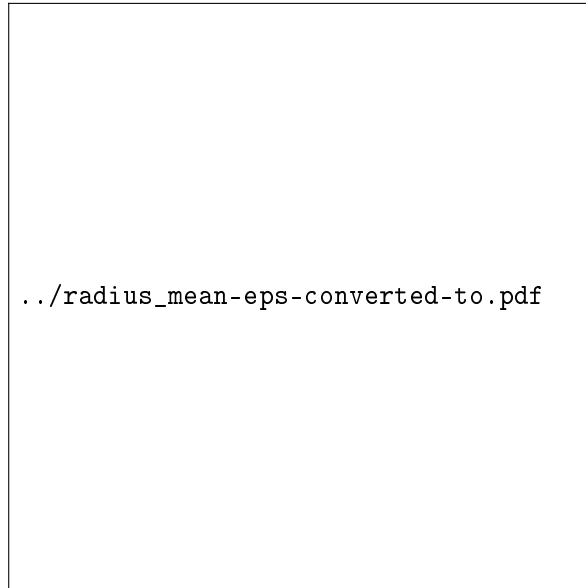


Figura 1: Mean

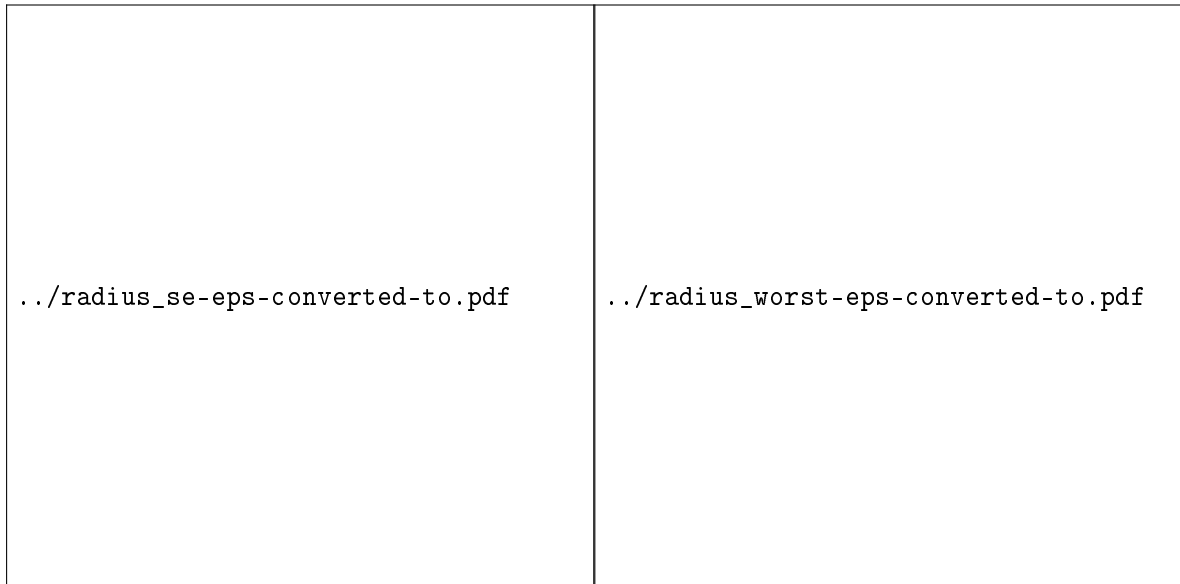


Figura 2: Standard Error

Figura 3: Worst

3.1 Classificador Bayesiano Simples

3.2 Classificador Bayesiano Quadrático

3.3 Mínimos Quadrados

3.4 Regressão Logística

4 Apresentação da Tecnologia