

Formação Inteligência Artificial

FEATURE SCALLING
DIMENSIONAMENTO DE CARACTERÍSTICAS

Dimensionamento de Características

Processo de transformação de dados numéricos

Variáveis em escalas diferentes

- Contribuem de forma desbalanceada para o modelo
- Exemplo: Salário e Altura

Gradient Descent converge mais rapidamente para o mínimo local



Padronização (Z-score)

- Dados aproximados da média (zero) e desvio padrão 1
- Podem ser negativos
- Não afeta outliers
- Deve ser usado na maioria dos casos

$$X_p = \frac{X - \mu}{\sigma}$$



Normalização (Min-Max)

- Transforma para escala comum entre zero e 1
- Usado em processamento de imagens e RNA
- Quando não sabemos a distribuição dos dados
- Quando precisam ser positivos
- Algoritmos não "requerem" dados normais
- Remove outliers pois impõe "limites"

$$X_n = \frac{X - X_{min}}{X_{max} - X_{min}}$$

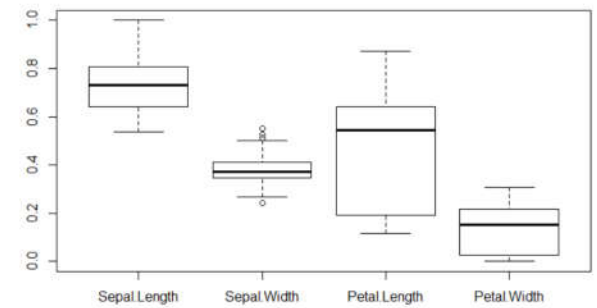
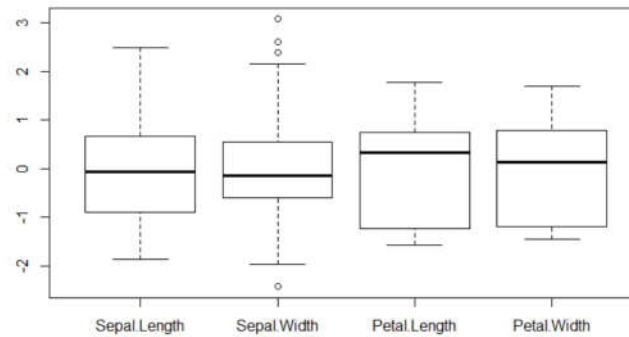
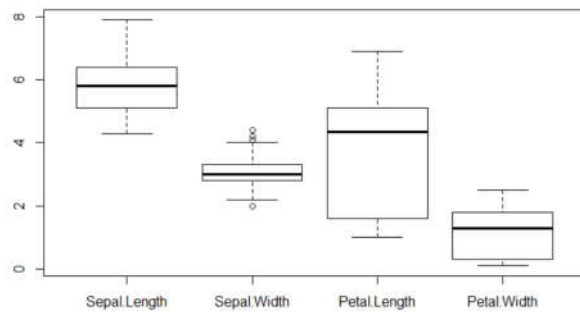


IRIS

IRIS

Padronização (Z-score)

Normalização (min-max)



Dimensionamento de Características

Não vai necessariamente melhorar seu modelo!

Arvores de decisão não precisam de nenhum tipo

Não se aplica a atributos categóricos transformados

