

Inteligência Computacional

Especificação do Projeto₁

Objetivo

Implementar uma solução de seleção de variáveis usando Wrapper. O algoritmo de busca deve ser o Algoritmo Genético. O algoritmo de aprendizagem deve ser o KNN com $K=3$.

Base de Dados

A base de dados para o projeto é AG.csv que está disponível no Github da disciplina. Essa base possui 40 atributos de entrada e 1 atributo alvo que é numérico, logo é um problema de regressão e deve ser usado o KNeighborsRegressor.

Etapas do Projeto

- Para simplificar o projeto, use a base inteira para selecionar as variáveis mais relevantes usando Wrapper com Algoritmo Genético e KNN;
- Use a métrica do Coeficiente de Determinação, também conhecido como R^2 , para avaliar os modelos. Esta métrica está disponível em `sklearn.metrics.r2_score`;
- Após selecionar as variáveis mais relevantes, gere uma nova base de dados apenas com as variáveis relevantes;
- Divida a nova base de dados em treinamento(70%) e teste (30%);
- Construa um modelo com o KNN com $k=3$ com a base de treinamento;
- Faça a predição do conjunto de teste com o modelo criado anteriormente;
- Calcule o `r2_score` para o conjunto de teste;
- Plot um gráfico de dispersão entre o Y previsto e o Y real do conjunto de teste;
- **Responda a pergunta:** O processo de seleção de variáveis melhorou a performance do modelo?