

Tarefa 1: Classificação e Regressão

Resgate de Vítimas de Catástrofes Naturais, Desastres ou Grandes Acidentes

Objetivo da tarefa

Construir regressores e classificadores e analisá-los para escolher o melhor dentre eles segundo as métricas de desempenho e de capacidade de generalização (viés e variância).

- Construir um classificador com CART ou com Redes Neurais MLP (RN): *dados os sinais vitais de 1 a 10, o classificador deve classificar a vítima de acordo com a característica 13 (tri).*
- Construir um regressor com CART ou com Redes Neurais MLP: *dados os sinais vitais de 1 a 10, o classificador deve classificar a vítima de acordo com a característica 14 (sobr).*
- **Obrigatoriamente** utilizar as duas técnicas. Por exemplo, se escolher CART para fazer o classificador, então escolha RNs para regressor e vice-versa.

Requisitos

1. INDIVIDUAL

2. O **formato do dataset** para treinamento está aqui.
3. Não é permitido utilizar as características que seguem no treinamento: 11 gcs e 12 avpu.
4. Gere datasets de treinamento/validação com o programa
https://github.com/tacla/VictSim3/blob/main/data_creation/gerar_dados_vitimas.py
O número de vítimas fica a seu critério assim com a divisão entre treino/validação.
5. Utilizar o **f-score** como medida de desempenho para os classificadores e **MSE NEGATIVO** para os regressores
6. Para cada preditor (classificador e regressor):
 - 6.1. escolher **três parametrizações** diferentes para treino/validação.
treinar o algoritmo com validação cruzada k-fold.
 - 6.2. comparar as três parametrizações e **escolher a melhor** (viés e variância; underfitting e overfitting).
 - 6.3. escolher um modelo dentro da parametrização escolhida segundo critérios explicitados por você.
 - 6.4. retrainar o **modelo escolhido** com todo o dataset
 - 6.5. realizar a predição com o modelo retrainado para o dataset que segue:
<https://github.com/tacla/VictSim3/tree/main/datasets/vict/1000v>

Observação:

Você pode se inspirar no código disponível no Google Colab

<https://colab.research.google.com/drive/1WBhNiUpY9gO3-u7jzQ1Z7mYFwkEM3Qx7?usp=sharing>

Formato do relatório

Apresentar na forma de tópicos:

- 1) Descrever os datasets utilizados, os parâmetros de criação de cada um deles e o número de folds da validação cruzada
- 2) Apresentar as três parametrizações (P1, P2 e P3) escolhidas para o **classificador**.
- 3) Apresentar os resultados de treino e validação no formato abaixo para cada parametrização i:

f-score	Mod. K=1	Mod. K=2	Mod. K=3	Média	Variância
treino	<valor>1	<valor>2	<valor>3		
Validação	<valor>1	<valor>2	<valor>3		

- 4) Comparar as médias dos f-scores das três parametrizações p1, p2 e p3:

MÉDIA f-score	P1	P2	P3
treino	<valor>1	<valor>2	<valor>3
Validação	<valor>1	<valor>2	<valor>3

- 5) Comparar os resultados da variância dos f-scores das três parametrizações p1, p2 e p3:

MÉDIA f-score	P1	P2	P3
treino	<valor>1	<valor>2	<valor>3
Validação	<valor>1	<valor>2	<valor>3

- 6) Escolher e justificar a escolha da melhor parametrização dentre P1, P2 e P3 com base nos critérios de under/overfitting (f-score e variância).
- 7) Escolher e justificar a escolha de um dos modelos da parametrização escolhida com base nos critérios que você vai definir.
- 8) Após realizar a predição com o modelo retreinado para o dataset de teste cego, apresente a matriz de confusão e os valores obtidos de precisão, recall e f-score.
- 9) Discuta e justifique com base nas tabelas de resultados, se o modelo escolhido apresenta boa capacidade de generalização.
- 10) Responda as perguntas de 2 a 9 para o **regressor**, utilizando **MSE NEGATIVO** no lugar de f-score. **Na pergunta 8**, apresente somente o valor de MSE NEGATIVO.

ENTREGA

- Somente no moodle
- Relatório em PDF
- Código