

Projeto da Disciplina – Aprendizado de Máquina: Algoritmos de Classificação em Linguagem C

Objetivo: a dupla deverá fazer uma implementação, em linguagem C, de um algoritmo de classificação. O algoritmo deve ser adaptado ao problema de classificação de células cancerígenas que envolve dados tabulares. Dados e descrição do dataset disponíveis [aqui](#).

Requisitos:

- Cada equipe irá escolher um algoritmo de classificação diferente.
- Do total de dados fornecidos na tarefa, 80% devem ser usados para treinamento dos modelos e 20% para a validação.
- Os critérios de avaliação de desempenho dos métodos serão os mesmos para todas as duplas: os 6 critérios apresentados nos próximos slides.

CrITÉRIOS de Avaliação de Desempenho do Método:

Serão considerados 6 critérios de avaliação, todos devem ser apresentados e avaliados (matriz de confusão e equações 1 a 5):

1. Matriz de confusão:

- TP: *true positives*;
- FP: *false positives*;
- TN: *true negatives*;
- FN: *false negatives*.

		Predicted class	
		P	N
Actual Class	P	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

2. Erro de Previsão (*Prediction Error - Err*):

$$Err = \frac{FP + FN}{FP + FN + TP + TN}$$

3. Acurácia (*Accuracy* - *Acc*):

$$Acc = \frac{TP + TN}{FP + FN + TP + TN}$$
$$Acc = 1 - Err$$

4. Precisão (*Precision* - *Pre*):

$$Pre = \frac{TP}{TP + FP}$$

5. Recall (*Rec*):

$$Rec = \frac{TP}{FP + FN}$$

6. *F1_score*:

$$F1score = \frac{2 \times Pre \times Rec}{Pre + Rec}$$

Entrega:

1. Código computacional que treina e avalia o método escolhido nos 6 critérios de avaliação de desempenho;
2. Apresentação em slides (7 min) contendo:
 - **Explicação do método (como ele funciona);**
Incluir autor e ano em que o método foi proposto.
Explicar como o método funciona.
 - **Procedimentos de treinamento e validação do método;**
Explicar como é feita a entrada de dados no método.
Valores de *Err*, *Acc*, *Pre*, *Rec*, *F1score* para conjunto de dados de treinamento.
 - **Resultados (critérios de avaliação);**
Valores de *Err*, *Acc*, *Pre*, *Rec*, *F1score* para o conjunto de dados de validação.
Incluir alguns detalhes de implementação.
 - **Discussão.**
Comentar se o método foi eficiente, como pode melhorar, o que foi feito de errado ou certo durante a implementação/treinamento/validação.

Observações:

- A equipe que procurar formas de melhorar a resposta dos métodos, poderá ter um diferencial na nota de até 1,0 ponto;
- Considerando as turmas de **Computação 2**, as equipes que apresentarem os melhores resultados serão convidadas pelos professores a escrever um artigo em conjunto para publicação em 2025.