

Tarefa 1_C ----- TEG

Mantenham as mesmas equipes da tarefa_1_A

Objetivo geral: avaliação dos agrupamentos obtidos por análise de componentes conexos sobre a tarefa1B usando algoritmos BFS ou DFS adaptados.

Linguagem de programação: C (exceto para a exibição de grafo via Python);

Objetivos específicos:

1. Avaliar o treinamento do clustering por meio de métricas usuais de Machine Learning;
4. Deseja-se avaliar o grau de sucesso na separação de espécies da flor Iris em duas classes.

Requisitos funcionais:

1. Partindo dos dois agrupamentos obtidos na tarefa1B, localize o agrupamento que tem população de maioria “setosa”, esta será a nomeação deste agrupamento, ficando o outro nomeado como grupo das “não-setosas”;
2. Com base na base de dados originais, conhecendo as classes corretas de cada flor, construa a matriz de confusão (veja a descrição abaixo);
3. Calcule a acurácia do procedimento de agrupamento (veja a descrição abaixo);

Observações:

1. Submissão da tarefa mediante relatório (veja no Moodle);
2. Programação em C (exceto para a exibição de grafo via Python);
3. Matriz de confusão/acurácia:
uma vez determinados os clusters, cada agrupamento poderá ser identificado com um tipo de flor (conforme descrito nos requisitos).

Ocorre que esse treinamento pode incorrer em erros, sendo necessário avaliar os resultados.

Por exemplo, um certo agrupamento pode conter todos os casos de setosa e possivelmente alguns casos erroneamente posicionados nesse grupo (Figura 1).

Com base na predição do modelo (agrupador) e na realidade conhecida (base de dados identificados) é possível construir uma matriz de confusão (veja a abaixo), identificando os casos TP (true positive), FP (false positive), TN (true negative) e FN (false negative) e realizar a extração das métricas de qualidade da classificação, por exemplo a acurácia:

Acurácia= $(TP+TN)/(TP+FP+TN+FN)$.

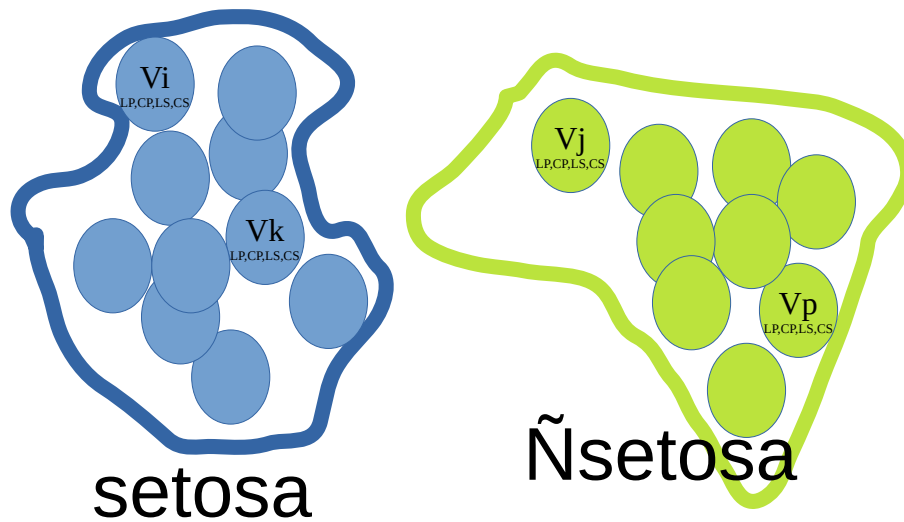
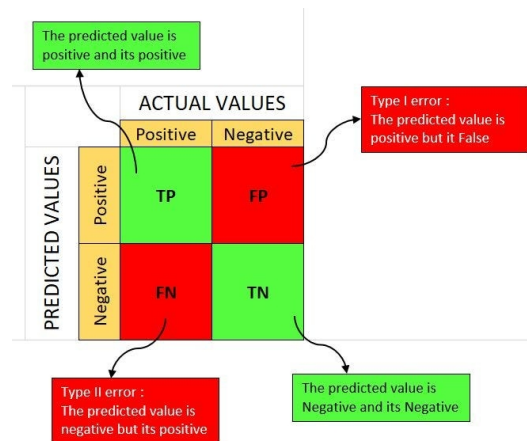


Figura 1: Resultado do treinamento do modelo proposto.

	sepal_length	sepal_width	petal_length	petal_width	variety
Vp	LS _p	WS _p	LP _p	LP _p	setosa
...					
Vk	LS _k	WS _k	LP _k	LP _k	virgínica
...					
Vi	LS _i	WS _i	LP _i	LP _i	setosa
...					
Vj	LS _j	WS _j	LP _j	LP _j	versicolor



“positivo” e “negativo” se referem à classe setosa			Real		
			setosa	versicolor	virgínica
			positivo	negativo	
Predito	setosa	positivo	TP	FP	
	versicolor	negativo	FN	TN	
	virgínica				