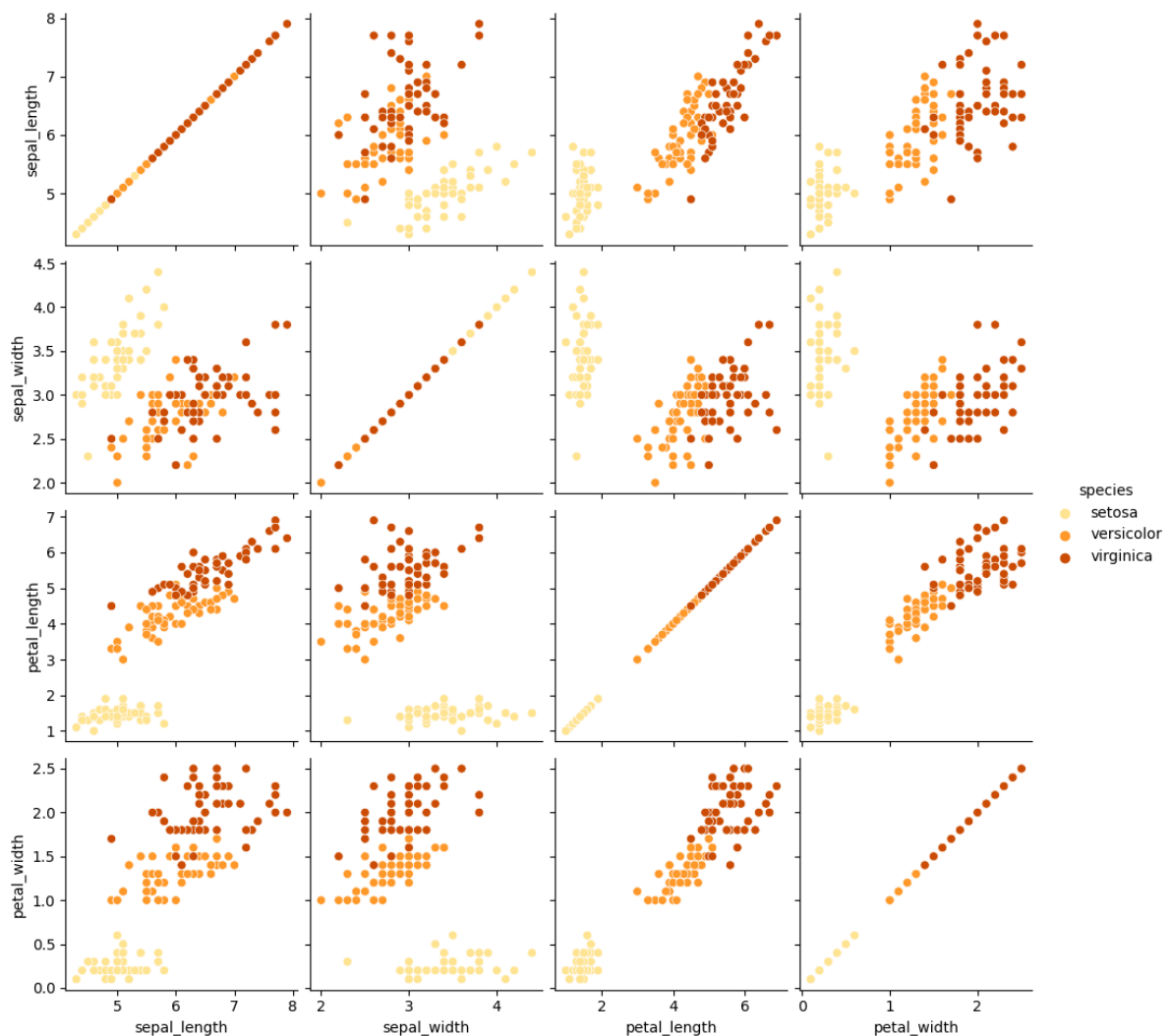


P5 - Resultados

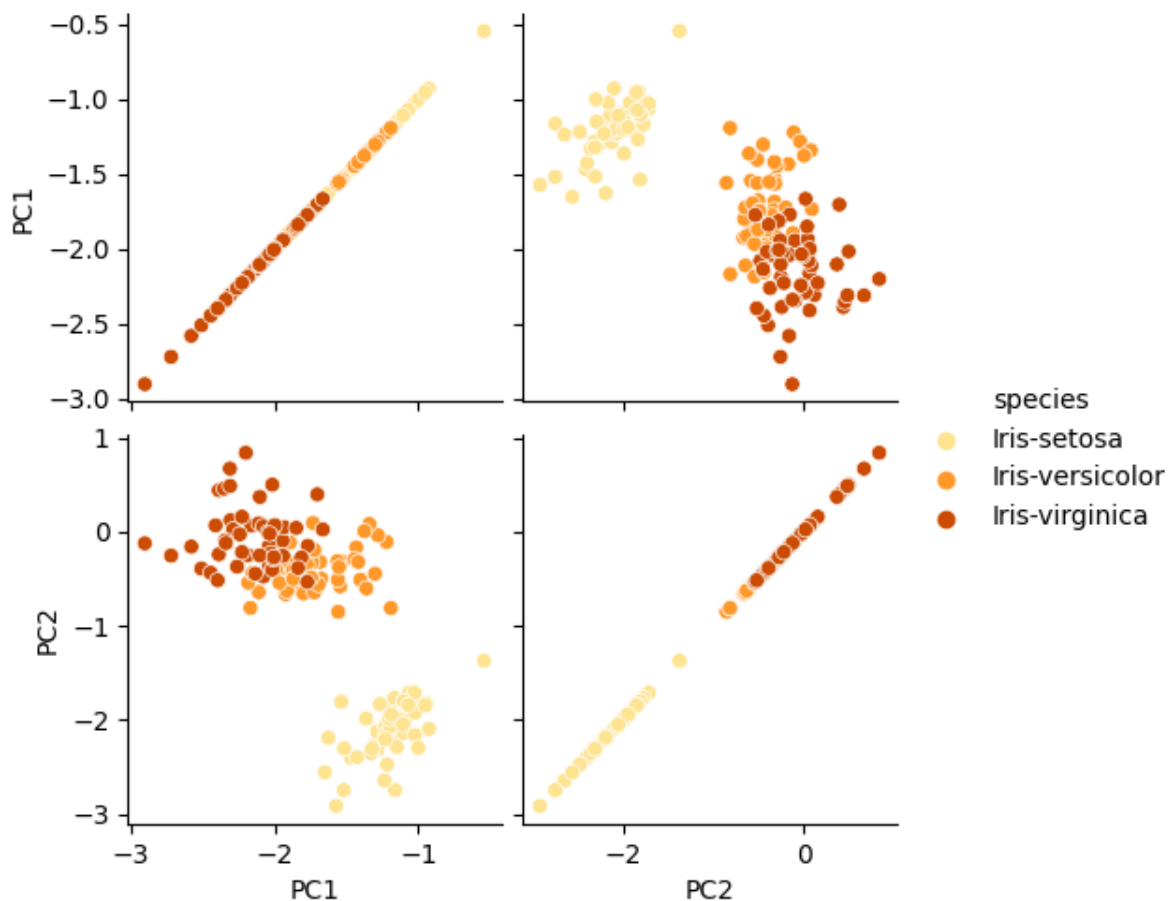
Parte A

Nesse projeto, iniciamos o procedimento com a leitura dos parâmetros referentes ao banco de dados Iris. Para testar tal leitura, graficamos os parâmetros de dois em dois com todas as variações possíveis, conseguindo a imagem a seguir.



Pode-se notar que cada um dos grupos de flores pode ser razoavelmente separado, dependendo do par de atributos escolhido, sendo que um deles (Iris-setosa) é geralmente mais diferenciado tendo em vista as outras duas espécies.

Em seguida, implementamos o PCA com base nesses quatro atributos, resultando em dois principais atributos mostrados da mesma forma na figura abaixo:



Vemos que esse método é de grande interesse pois abstrai os parâmetros iniciais (4) em dois que carregam informações suficientes para que os grupos sejam separados de forma similar, porém agora com menos dados. Vemos ainda que a característica de separação da espécie Iris-setosa ainda é mais bem definida em detrimento das outras duas.

Com os dados do PCA em mãos, fomos capazes de separar nosso conjunto de plantas em duas categorias, uma para teste e outra para o treinamento, com o intuito de inferirmos a categoria das plantas do conjunto de teste comparando-as com todas as do conjunto de treinamento pelo método dos K-vizinhos. Nele, calculamos a distância entre nosso estímulo (do conjunto de teste) e os pontos de apoio (plantas do conjunto de treinamento), inferindo a categoria pela menor distância encontrada. Dessa forma, construímos a seguinte matriz de confusão:

		WithPCA		
		Iris-versicolor	Iris-virginica	Iris-setosa
Inferido	Iris-versicolor	15	6	0
	Iris-virginica	7	22	0
	Iris-setosa	0	0	25
		Real		

A implementação desse método se mostra satisfatória pois diminui a complexidade de nossa análise e mantém um ótimo desempenho quanto a inferência da espécie da planta de estímulo.