Identificar el minimo numero de variables

1. COVARIANZA
2. PCA
3. ANALISIS DE COVARIANZA

Se puede encontrar que componentes son las representativos, al eliminar los componentes que son altamente correlacionados con otros, y luego aplicar algebra lineal.

Si llega a la misma conclusion si usa COVARIANZA o PCA, la ventaja de COVARIANZA es hacer una exploracion de correlaciones entre componentes.

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.preprocessing import StandardScaler

iris = sns.load\_dataset('iris')

scaler = StandardScaler()

scaled = scaler.fit\_transform(

          iris[['sepal\_length', 'sepal\_width','petal\_length','petal\_width']].values

            )

covariance\_matrix = np.cov(scaled.T)

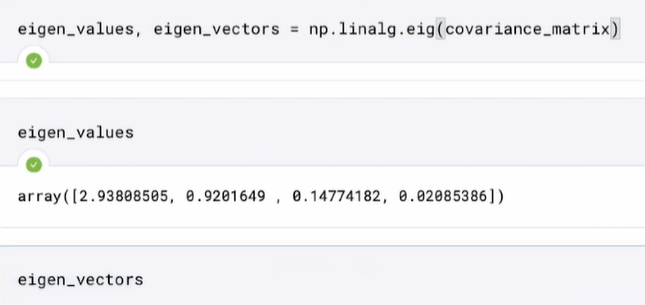
sns.pairplot(iris)

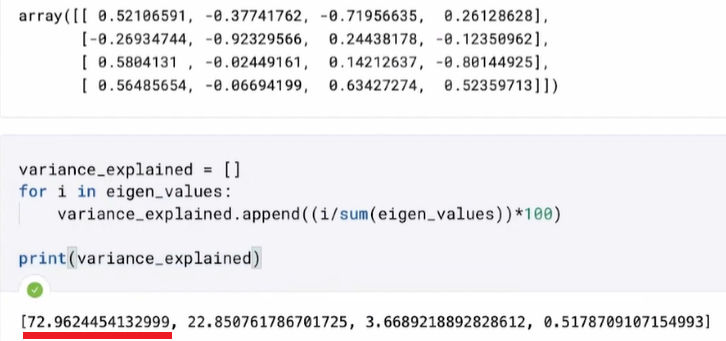
sns.pairplot(iris , hue='species')

|  |  |
| --- | --- |
| sns.pairplot(iris) | sns.pairplot(iris , hue=’species’) |
| Chart, histogram  Description automatically generated |  |



|  |  |
| --- | --- |
|  |  |





LO QUE HEMOS VISMO ANTES CON VARIANZA Y VECTORES SE RESUME USANDO LO SIGUIENTE,

LLEGAMOS A LA MISMA CONCLUSION QUE 2 COMPONENTES SONLOS PRINCIPALES, QUE APORTAN 72.9% Y 22.85%

1. ANALISIS DE COVARIANZA

from sklearn.decomposition import PCA

pca =PCA(n\_components=2)

pca.fit(scaled)

pca.explained\_variance\_ratio\_

#0.7296 0.2284