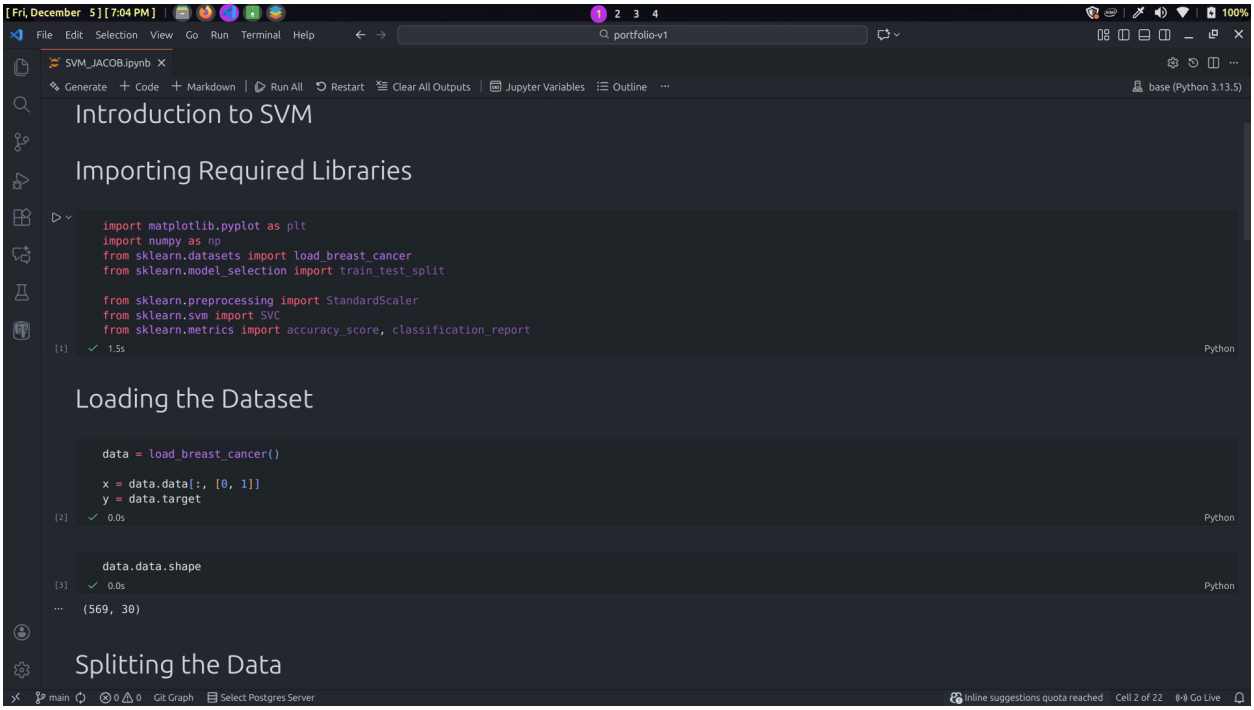
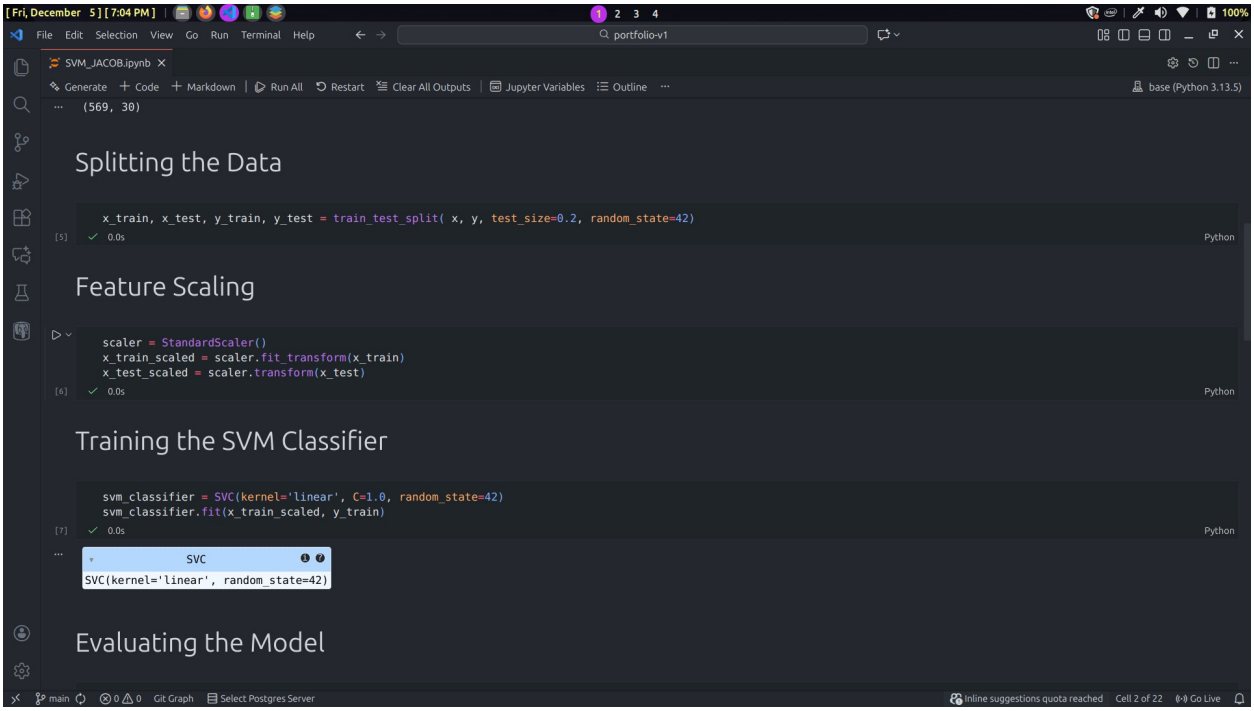
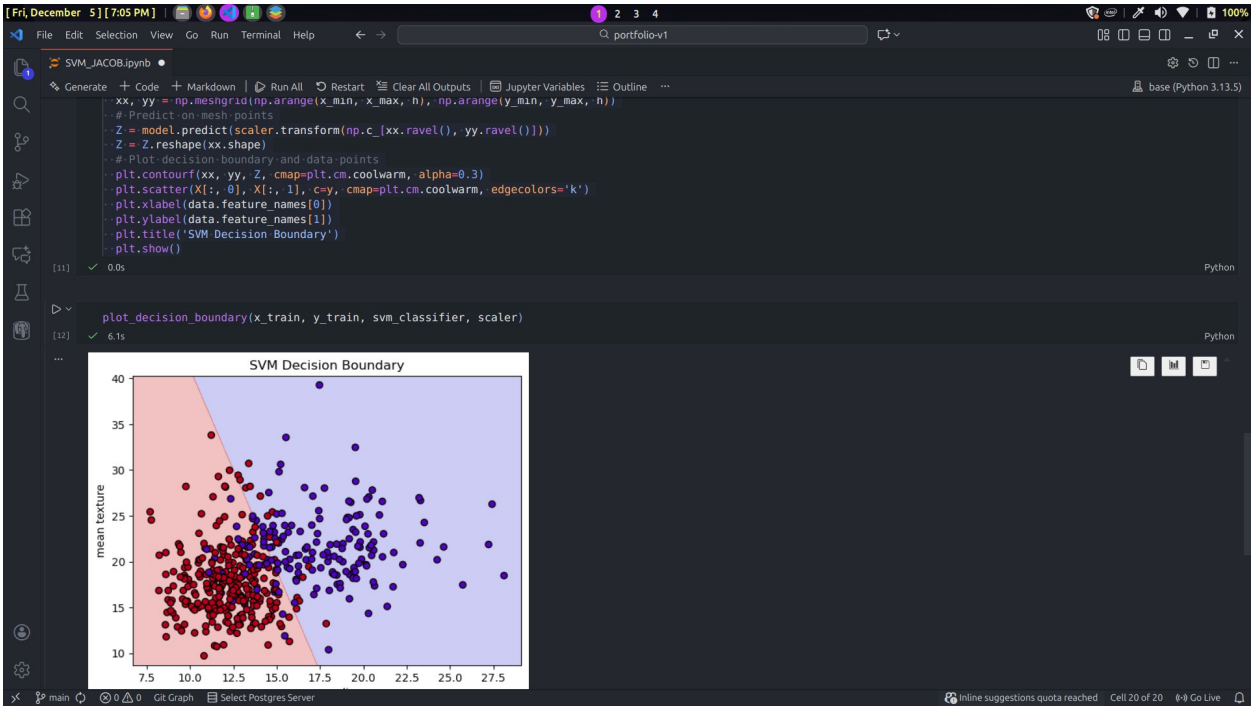

SVM_JACOB.ipynb
Introduction to SVM
Importing Required Libraries
import matplotlib.pyplot as plt
import numpy as np
from sklearn.datasets import load_breast_cancer
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score, classification_report
Loading the Dataset
data = load_breast_cancer()
x = data.data[:, [0, 1]]
y = data.target
data.data.shape
(569, 30)
Splitting the Data


SVM_JACOB.ipynb
(569, 30)
Splitting the Data
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=42)
Feature Scaling
scaler = StandardScaler()
x_train_scaled = scaler.fit_transform(x_train)
x_test_scaled = scaler.transform(x_test)
Training the SVM Classifier
svm_classifier = SVC(kernel='linear', C=1.0, random_state=42)
svm_classifier.fit(x_train_scaled, y_train)
SVC
SVC(kernel='linear', random_state=42)
Evaluating the Model



Interpretation of Results

The SVM classification plot shows a binary classification problem using mean radius and mean texture as features. The linear decision boundary does a pretty good job of separating the two classes, with red points on one side and blue points on the other.

It looks like mean radius is the most important feature because red points mostly appear in the lower range (7–15 units), while blue points spread across a higher range (12–28 units). Mean texture overlaps a lot between the classes (10–40 units), so it seems less helpful for distinguishing them. The red class is more tightly grouped, whereas the blue class is more spread out and varied.

Some points are on the “wrong” side of the boundary, showing that the classes aren’t perfectly linearly separable. This is especially noticeable in the middle region where red and blue points mix, making classification harder. These misclassified points are more about the data itself than a problem with the SVM.

Conclusion

The SVM does a decent job classifying the points with a simple linear boundary. But the overlap between classes and the misclassifications suggest that the relationship between the features and class isn’t fully linear. Mean radius is clearly the most important feature, but the imperfect separation means that adding more features or using a non-linear kernel (like RBF or

polynomial) could improve results. Overall, this shows that a linear SVM is simple and easy to interpret, but real-world data often has some unavoidable ambiguity.