Lab3_SupportVectorMachine_Report

1. **What's the difference between the different SVM kernels?**

   An SVM maps inputs into a (possibly higher-dimensional) feature space and finds a hyperplane that best separates classes (maximizes margin). The **kernel** function defines how input points are compared (i.e., the inner product in some feature space). Different kernels correspond to different feature mappings and therefore different shapes of decision boundary.

   Linear Kernel

   Decision Boundary Shape: Straight line / hyperplane

   Hyperparameters: C

   Good For: Linearly separable or high-dimensional sparse data (e.g. text)

   Output score (decision_function): Signed distance to the hyperplane

   Polynomial Kernel

   Decision Boundary Shape: Polynomial curve (2nd, 3rd… degree)

   Hyperparameters: degree, coef0, gamma, C

   Good For: Data with feature **interactions** or curved relationships

   Output score (decision_function): Distance to polynomial decision surface

   RBF Kernel

   Decision Boundary Shape: Highly flexible, smooth nonlinear boundary

   Hyperparameters: gamma, C

   Good For: General-purpose nonlinear problems where the structure is unknown

   Output score (decision_function): Distance to nonlinear boundary in infinite-dimensional space

2. **When would you use each one, and how do their outputs differ?**

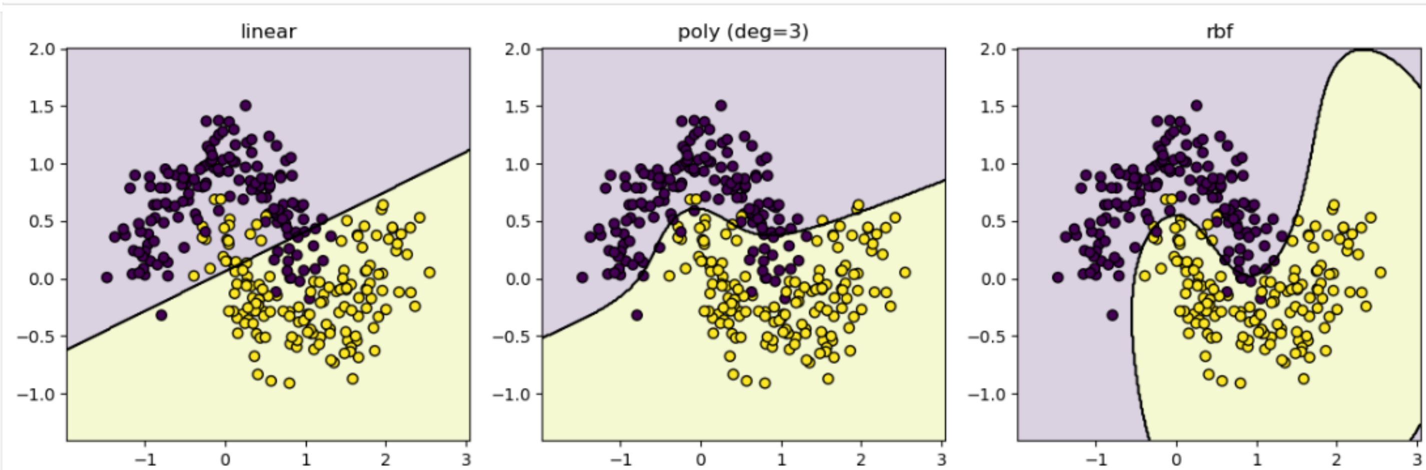
   Linear kernel (kernel='linear')
   - When the classes are **separable or approximately separable by a linear boundary**.
   - Also ideal when you have **very high dimensional features**, even if data is noisy.
   - Very fast for large datasets (use LinearSVC if you have a lot of samples).

   Polynomial kernel (kernel='poly', degree=d)

   - When you suspect polynomial relationships or interactions among features (e.g., quadratic or cubic effects).
   - Small- to medium-sized problems (higher degree increases complexity).
   - Degree 2 or 3 is common (higher → risk of overfitting).

   Radial Basis Function (RBF) kernel (kernel='rbf') — most commonly used
   - Good default when you don't know the right kernel — it's a general-purpose nonlinear kernel.
   - It can model **very complex nonlinear structure**.
   - Usually gives the best performance after proper tuning (C and gamma).
   - Good for most small/medium tabular datasets.

As per the output, Higer ROC-AUC (better separation) between the classes in the decision_function output.
The RBF model produces decision scores almost positives get higher scores than negatives, it is the highest among the three.
Polynomial kernel has the lowest ROC-AUC, some negatives are assigned high scores and some positives low scores (lower ranking ability).
Liner is in the middle, it separates reasonably well, but not as cleanly as RBF.



Linear SVM classification report

```
Linear SVM classification report
Consufion Matrix:  [[8299, 242], [16, 121]]
Classification Report
              precision    recall  f1-score   support

           0       1.00      0.97      0.98      8541
           1       0.33      0.88      0.48       137

    accuracy                           0.97      8678
   macro avg       0.67      0.93      0.73      8678
weighted avg       0.99      0.97      0.98      8678

Roc AUC Score:  0.9515356156692024
```

Poly SVM classification report

```
Poly SVM classification report
Consufion Matrix:  [[8485, 56], [28, 109]]
Classification Report
              precision    recall  f1-score   support

           0       1.00      0.99      1.00      8541
           1       0.66      0.80      0.72       137

    accuracy                           0.99      8678
   macro avg       0.83      0.89      0.86      8678
weighted avg       0.99      0.99      0.99      8678

Roc AUC Score:  0.9211933507503949
```

RBF SVM classification report

```
RBF SVM classification report
Consufion Matrix:  [[8450, 91], [31, 106]]
Classification Report
              precision    recall  f1-score   support

           0       1.00      0.99      0.99      8541
           1       0.54      0.77      0.63       137

    accuracy                           0.99      8678
   macro avg       0.77      0.88      0.81      8678
weighted avg       0.99      0.99      0.99      8678

Roc AUC Score:  0.972283113568985
```