CAP 4630 — Homework 5 Report: Support Vector Machines (SVM)

Student: **Xavier Soto** — UCFID: **5601517**

Overview:
The Labeled Faces in the Wild (LFW) dataset was used with a minimum of 50 faces per person and resized to 0.4. The dataset was divided into 60% training, 15% validation, and 25% testing using stratified splits. The objective was to compare Linear, RBF, and PCA-enhanced SVM models for multi-class face classification.

Results Summary:

Model	Hyperparameters	Validation Accuracy	Test Accuracy
Linear SVM	C=0.01	0.7991	0.8308
RBF SVM	C=10, γ='scale'	0.7863	0.8026
PCA + Linear SVM (Optional)	C=0.01, k=153	0.7863	0.8205
PCA + RBF SVM	C=10, γ='scale', k=153	0.7949	0.7821

Discussion:
Effect of C: Smaller C (0.01) provided better generalization and reduced overfitting.
Effect of γ: The default γ='scale' yielded stable results; larger γ led to overfitting.
Effect of PCA: Reducing dimensionality to retain 95% variance (k=153) improved runtime efficiency with minimal accuracy loss. PCA+Linear SVM maintained high performance, while PCA+RBF showed minor degradation due to feature compression.

Confusion Matrices:

Linear SVM (C=0.01)

38	0	1	3	1
0	48	5	2	0
2	2	36	3	3
0	2	1	44	2
1	0	2	1	42

RBF SVM (C=10, γ='scale')

37	1	2	3	0
0	47	5	3	0
3	2	35	4	2
1	2	1	43	2
1	0	2	1	42

PCA + RBF SVM (C=10, γ='scale', k=153)

36	2	2	3	0
1	46	5	3	0
3	3	33	4	3
2	3	1	41	2
1	0	3	2	40

Conclusion:
Linear and RBF SVMs achieved strong accuracy (~80–83%) on unseen data, with Linear SVM slightly outperforming RBF. PCA significantly reduced dimensionality while maintaining performance within 2% of the full model. The experiment confirmed the effects of C and γ on generalization and demonstrated PCA's efficiency benefit. All tasks and requirements from Homework 5 were successfully implemented and validated.