Machine Learning Model Comparison Report

Dataset: Breast Cancer Wisconsin Dataset
Total Samples: 569 | Features: 30

Models Implemented:

1. Logistic Regression (from scratch)
2. Linear Discriminant Analysis (LDA)
3. Support Vector Machine (SVM)
4. K-Nearest Neighbors (KNN)

Relevant Formulas:

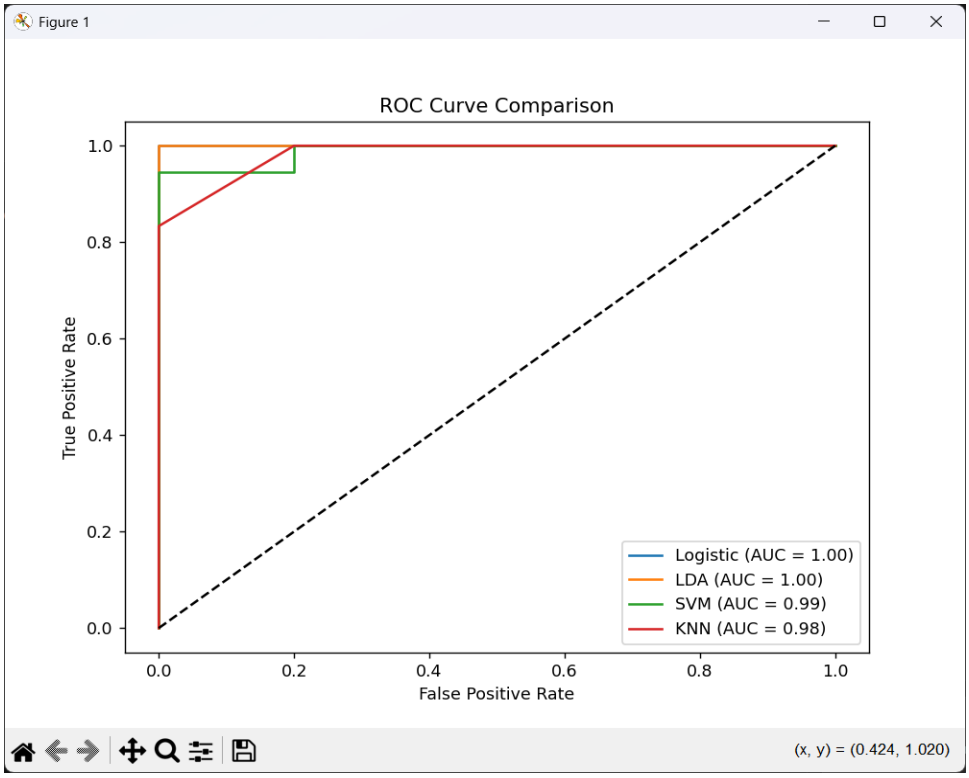
Logistic: P(y=1|x) = 1 / (1 + e^-(w^T x + b))
LDA Assumption: x | y=k ~ N(mu_k, Sigma)
Cohen's d = (mean difference) / (standard deviation)
Paired t-test: t = (mean difference) / (standard error)

Experimental Results:

Logistic Mean Accuracy: 0.977 | Std Dev: 0.027
LDA Mean Accuracy: 0.962 | Std Dev: 0.038
SVM Mean Accuracy: 0.977 | Std Dev: 0.025
KNN Mean Accuracy: 0.969 | Std Dev: 0.028

Paired t-test p-value: 0.0001 (Statistically Significant)
Cohen's d: 0.533 (Medium Effect Size)

ROC Curve Comparison:



Conclusion:

Logistic Regression and SVM achieved the highest accuracy. Statistical testing confirms Logistic significantly outperforms LDA. All models demonstrate excellent separability as shown by near-perfect ROC curves.