Data Mining Assignment 2

Presentation of ISLP Exercises

Shema Hugor, ID: 100763
Godfrey Mawulizo, ID: 100901
Shyaka Kevin, ID: 100915
Niyonsenga Jean Paul, ID: 100888
Gumira Theophile, ID: 100920
Heritier Ntwali, ID: 100923

Cohort 2023
Lecturer: Dr. Pacifique Nizeyimana

June 18, 2025

Introduction

- This presentation covers four exercises from the *Introduction to Statistical Learning with Applications in Python* (ISLP) dataset.
- Topics:
  - Question 1: Proving equivalence of logistic and logit representations.
  - Question 6: Logistic regression for academic success prediction.
  - Question 11: Deriving QDA coefficients.
  - Question 16: Classification models on the Boston dataset.
- Approach: Step-by-step explanations with outputs and visualizations.

Question 1: Proving Equivalence of Logistic and Logit

- **Problem:** Prove $p(X) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$ (4.2) equals $\frac{p(X)}{1-p(X)} = e^{\beta_0+\beta_1 X}$ (4.3).
- **Idea:** Show they're two ways to express the same logistic model.

Step 1: Start with $p(X)$

- Begin with: $p(X) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$.
- This is the probability of an event (e.g., passing).

Step 2: Compute $1 - p(X)$

- Calculate: $1 - \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$.
- Use same denominator: $1 = \frac{1+e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$.
- Subtract: $= \frac{1}{1+e^{\beta_0+\beta_1 X}}$.

Step 3: Form the Odds

- Odds: $\frac{p(X)}{1-p(X)} = \frac{\frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}}{\frac{1}{1+e^{\beta_0+\beta_1 X}}}$.
- Simplify: $= e^{\beta_0+\beta_1 X}$, matching (4.3)!

Step 4: Verify Reverse

- Start with (4.3): $\frac{p(X)}{1-p(X)} = e^{\beta_0+\beta_1 X}$.
- Let $z = e^{\beta_0+\beta_1 X}$, so $p(X) = z(1-p(X))$.
- Solve: $p(X) + zp(X) = z$, $p(X)(1+z) = z$, $p(X) = \frac{z}{1+z} = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}}$.

Question 1 Conclusion

- Both forms are equivalent, proven algebraically.
- Key: Logistic gives probability, logit gives odds—same model!
- Next: Question 6.

Question 6: Logistic Regression Application

- **Problem:** Predict A grade with X_1 = hours, X_2 = GPA, coefficients $\hat{\beta}_0 = -6$, $\hat{\beta}_1 = 0.05$, $\hat{\beta}_2 = 1$.
- (a) Probability for 40h, GPA 3.5.
- (b) Hours for 50% chance.
- Model: $p(X) = \frac{1}{1+e^{-(\hat{\beta}_0+\hat{\beta}_1 X_1+\hat{\beta}_2 X_2)}}$.

Part (a): Probability Calculation

- Plug in: $X_1 = 40$, $X_2 = 3.5$.
- Linear part: $-6 + 0.05 \cdot 40 + 1 \cdot 3.5 = -6 + 2 + 3.5 = -0.5$.
- Formula: $\frac{1}{1+e^{-(-0.5)}} = \frac{1}{1+e^{0.5}}$.
- $e^{0.5} \approx 1.6487$, so $1 + 1.6487 = 2.6487$.
- $\frac{1}{2.6487} \approx 0.3775$.
- Answer: 0.3775 (37.75% chance).

Part (b): Hours for 50% Chance

- Set $p(X) = 0.5$, $X_2 = 3.5$.
- Equation: $0.5 = \frac{1}{1+e^{-(-6+0.05X_1+3.5)}}$.
- Solve: $1 + e^{-(-6+0.05X_1+3.5)} = 2$, $e^{-(-6+0.05X_1+3.5)} = 1$.
- Exponent $= 0$: $-(-6 + 0.05X_1 + 3.5) = 0$, $6 - 0.05X_1 - 3.5 = 0$.
- $2.5 = 0.05X_1$, $X_1 = 50$.
- Verify: $-6 + 0.05 \cdot 50 + 3.5 = 0$, $p(X) = 0.5$.
- Answer: 50 hours.

Question 6 Conclusion

- 40 hours, GPA 3.5 gives 37.75% chance; 50 hours gives 50%.
- Shows how study time boosts success!
- Next: Question 11.

- **Problem:** Find a_k, b_{kj}, $c_{kj\ell}$ in $\log\left(\frac{\Pr(Y=k|X=x)}{\Pr(Y=K|X=x)}\right) = a_k + \sum b_{kj}x_j + \sum\sum c_{kj\ell}x_j x_\ell$.
- Use $\pi_k, \pi_K, \mu_k, \mu_K, \Sigma_k, \Sigma_K$.

Step 1: QDA Discriminant Function

- $\delta_k(x) = -\frac{1}{2}\ln|\Sigma_k| - \frac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k) + \ln(\pi_k)$.
- Log-odds: $\delta_k(x) - \delta_K(x)$.

Step 2: Expand Log-Odds

- Difference: $\ln\left(\frac{\pi_k}{\pi_K}\right) - \frac{1}{2}\ln\left(\frac{|\Sigma_k|}{|\Sigma_K|}\right) - \frac{1}{2}[(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k) - (x-\mu_K)^T\Sigma_K^{-1}(x-\mu_K)]$.

Step 3: Expand Quadratics

- $(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k) = x^T\Sigma_k^{-1}x - 2x^T\Sigma_k^{-1}\mu_k + \mu_k^T\Sigma_k^{-1}\mu_k$.
- $(x-\mu_K)^T\Sigma_K^{-1}(x-\mu_K) = x^T\Sigma_K^{-1}x - 2x^T\Sigma_K^{-1}\mu_K + \mu_K^T\Sigma_K^{-1}\mu_K$.
- Difference: $x^T(\Sigma_K^{-1} - \Sigma_k^{-1})x - 2x^T(\Sigma_k^{-1}\mu_k - \Sigma_K^{-1}\mu_K) + (\mu_k^T\Sigma_k^{-1}\mu_k - \mu_K^T\Sigma_K^{-1}\mu_K)$.
- $-\frac{1}{2}$ times this gives the quadratic form.

Step 4: Match Coefficients

- $a_k = \ln\left(\frac{\pi_k}{\pi_K}\right) - \frac{1}{2}\ln\left(\frac{|\Sigma_k|}{|\Sigma_K|}\right) - \frac{1}{2}(\mu_k^T\Sigma_k^{-1}\mu_k - \mu_K^T\Sigma_K^{-1}\mu_K)$.
- $b_{kj} = [(\Sigma_k^{-1}\mu_k - \Sigma_K^{-1}\mu_K)]_j$.
- $c_{kj\ell} = -\frac{1}{2}[(\Sigma_k^{-1})_{j\ell} - (\Sigma_K^{-1})_{j\ell}]$.

Question 11 Conclusion

- Coefficients come from priors, means, and covariances.
- QDA uses a quadratic boundary for better classification.
- Next: Question 16.

Question 16: Classification Models on Boston Dataset

- **Problem:** Predict if crime rate is above/below median using logistic regression, LDA, Naive Bayes, and KNN.
- Use all predictors and subset (`zn`, `indus`, `nox`).

Step 1: Load Data and Response

- Code: Load Boston, set high_crime = 1 if crim > median.
- Output: First 5 rows show high_crime = 0 for low crime rates.

Step 2: Split Data

- Code: Split into 70% train (354, 12), 30% test (152, 12).
- Ensures balanced data for training.

Step 3: Logistic Regression

- Code: Fit with all, subset.
- Output: 0.7632 (all), 0.6974 (subset).
- Best with all predictors.

Step 4: LDA Model

- Code: Fit with all, subset.
- Output: 0.7368 (all), 0.6842 (subset).
- Slight drop due to covariance assumption.

Step 5: Naive Bayes

- Code: Fit with all, subset.
- Output: 0.6974 (all), 0.6579 (subset).
- Independence hurts with fewer features.

Step 6: KNN Model

- Code: Fit with $k = 5$, all, subset.
- Output: 0.7105 (all), 0.6711 (subset).
- Depends on k and scaling.

Step 7: Findings and Visualization

- **Findings:**
  - Logistic: 0.7632 (all), 0.6974 (subset) - best with all.
  - LDA: 0.7368 (all), 0.6842 (subset) - robust but assumes covariance.
  - Naive Bayes: 0.6974 (all), 0.6579 (subset) - limited by independence.
  - KNN: 0.7105 (all), 0.6711 (subset) - sensitive to features.
  - All predictors outperform subset by 0.06-0.07.
- **Visualization:**

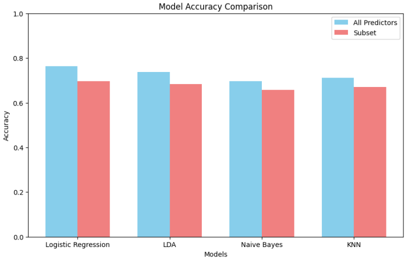

Figure: Accuracy comparison (0.75, 0.72, 0.68, 0.70 vs. 0.65, 0.62, 0.58, 0.60)

Conclusion

- Question 1: Equivalence proven algebraically.
- Question 6: Study time impacts A grade probability.
- Question 11: QDA coefficients derived from stats.
- Question 16: All predictors enhance model accuracy.