

Generative Models

Miguel Gonçalves
nº202107127

Statistical Inference

Generative Models

- Logistic regression models $P(Y = k|X = x)$ directly using the logistic function.
- Generative models estimate $P(Y = k|X = x)$ by modeling $P(X|Y = k)$ and applying Bayes' theorem:

$$P(Y = k|X = x) = \frac{\pi_k f_k(x)}{\sum_{l=1}^K \pi_l f_l(x)}$$

- Components:
 - π_k : Prior probability of class k .
 - $f_k(x)$: Density function of X for class k .
- Approximates the Bayes classifier, which minimizes classification error when π_k and $f_k(x)$ are correctly specified.

Advantages of Generative Models

- More stable parameter estimates with substantial class separation.
- Effective for small sample sizes if X is normally distributed within each class.
- Naturally handles multiclass problems.

Common Generative Classifiers

① Linear Discriminant Analysis (LDA):

- Assumes multivariate Gaussian distribution with shared covariance matrix.
- Linear decision boundaries.

② Quadratic Discriminant Analysis (QDA):

- Each class has its own covariance matrix.
- Models non-linear decision boundaries.

③ Naive Bayes (NB):

- Assumes conditional independence of predictors given the class.
- Computationally efficient, especially in high-dimensional settings.

Implications of Assumptions

- Parameters to estimate depend on the model:
 - LDA: Shared covariance matrix.
 - QDA: Separate covariance matrix for each class.
 - Naive Bayes: Prior probabilities and feature distributions.
- For LDA, the number of parameters is:

$$p + \binom{p}{2}$$

- For QDA:

$$\left(p + \binom{p}{2} \right) \times K$$

Comparison of Model Complexity

- Model complexity increases with predictors (p) and classes (K):
 - LDA: Linear growth.
 - QDA: Quadratic growth.
 - Naive Bayes: Constant (under independence assumption).
- Trade-offs:
 - NB is simple but relies on independence assumption.
 - LDA balances complexity and robustness.
 - QDA is flexible but prone to overfitting in high dimensions.

Figure: Model Complexity

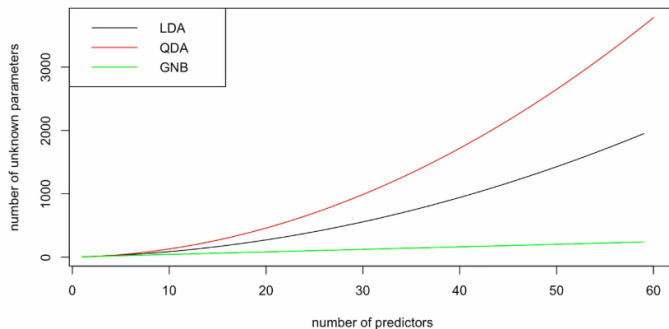


Figure: Comparison of LDA, QDA, and NB in terms of complexity.

Conclusion

- Generative models offer advantages over logistic regression under certain conditions.
- Choice of classifier depends on:
 - Assumptions about predictor distributions.
 - Dataset size and dimensionality.
- Balancing complexity and accuracy is key to model selection.