MA477: Data Science

Lesson 13 Outline — 10 February 2026

United States Military Academy, West Point
Instructor: MAJ Patrick Kuiper

---

1 Administrative

- Student review
- LDA Questions
- LDA Lecture

2 Linear Discriminate Analysis Lesson Objectives

- Understand the difference between discriminative models and generative models.
- Derive the discriminant function for Linear Discriminant Analysis (LDA) with a single predictor.
- Use sklearn modules to assess LDA classifiers.

3 Student Review

Discussion Questions: Bayes' Rule, LDA, and Logistic Regression

Question 1: What role does Bayes' rule play in Linear Discriminant Analysis (LDA)?

Answer:

Bayes' rule provides the mechanism by which Linear Discriminant Analysis (LDA) converts a model of the data-generating process into a classifier. Specifically, LDA models the class-conditional distribution

$$f_k(x) = \Pr(X = x \mid Y = k),$$

which represents the probability density of observing predictor values x given that the observation belongs to class k, as well as the class prior

$$\pi_k = \Pr(Y = k),$$

which represents the probability that a randomly selected observation belongs to class k before observing any predictors.

Bayes' rule combines these quantities to compute the posterior class probability. In general, Bayes' rule can be written as

$$\Pr(Y = k \mid X = x) = \frac{\Pr(X = x \mid Y = k)\Pr(Y = k)}{\Pr(X = x)}.$$

In the context of classification with K classes, the denominator $\Pr(X = x)$ can be expressed as a sum over all classes, yielding

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

Here, X denotes the predictor vector, Y denotes the class label, $f_k(x)$ is the class-conditional density for class k, π_k is the prior probability of class k, and the summation in the denominator ensures that the posterior probabilities across all classes sum to one.

The LDA classifier assigns an observation x to the class with the largest posterior probability. Thus, Bayes' rule is essential for transforming LDA from a density estimation problem into a classification method.

—

Question 2: What assumptions does LDA make about the distribution of the predictors, and why are these assumptions important?

Answer: LDA assumes that the predictor vector $X \in \mathbb{R}^p$ follows a multivariate Gaussian distribution within each class:

$$X \mid (Y = k) \sim \mathcal{N}(\mu_k, \Sigma),$$

where each class has its own mean vector μ_k, but all classes share a common covariance matrix Σ. These assumptions are important because they lead to a closed-form expression for the posterior probabilities and result in linear decision boundaries. The shared covariance assumption causes quadratic terms in x to cancel when comparing classes, yielding discriminant functions that are linear in x.

—

Question 3: Why is Linear Discriminant Analysis considered a generative model?

Answer: LDA is considered a generative model because it explicitly models the joint distribution of the data and the class labels:

$$\Pr(X, Y) = \Pr(X \mid Y) \Pr(Y).$$

By modeling $\Pr(X \mid Y = k)$ for each class and the class prior $\Pr(Y = k)$, LDA describes how observations are generated. In principle, one could generate synthetic data by first sampling a class label from $\Pr(Y)$ and then sampling feature values from $\Pr(X \mid Y)$. Classification is then performed by applying Bayes' rule to reverse this generative process and compute $\Pr(Y \mid X)$.

—

Question 4: How does logistic regression differ from LDA from a probabilistic modeling perspective?

Answer: Logistic regression is a discriminative model because it directly models the conditional probability $\Pr(Y \mid X)$ without modeling the distribution of X. In the binary case, logistic regression assumes

$$\log\left(\frac{\Pr(Y = 1 \mid X = x)}{\Pr(Y = 0 \mid X = x)}\right) = \beta_0 + \beta^\top x.$$

In contrast, LDA models $\Pr(X \mid Y)$ and $\Pr(Y)$ separately and then uses Bayes' rule to compute $\Pr(Y \mid X)$. Logistic regression makes fewer assumptions about the distribution of X, whereas LDA relies on Gaussian assumptions for interpretability and stability.

—

Question 5: Why does LDA produce linear decision boundaries, and how does this relate to logistic regression?

Answer: LDA produces linear decision boundaries because it assumes a common covariance matrix Σ across all classes. When the Gaussian class-conditional densities are substituted into Bayes' rule, the resulting discriminant function for class k takes the form

$$\delta_k(x) = x^\top \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^\top \Sigma^{-1} \mu_k + \log \pi_k,$$

which is linear in x. The decision boundary between two classes is therefore defined by a linear equation. This is closely related to logistic regression, which also produces linear decision boundaries through a linear model for the log-odds. Under the LDA assumptions, the posterior probabilities from LDA and logistic regression can be very similar.