Solution 1:

(a) **Assumptions of Each Method:**

- **Quadratic Discriminant Analysis (QDA):**
  - Each class has its own class-specific mean vector $\boldsymbol{\mu}_k$ and covariance matrix Σ_k.
  - Does not assume feature independence.
  - Decision boundaries are quadratic as covariance matrices can differ between classes.
- **Linear Discriminant Analysis (LDA):**
  - Special case of QDA with different mean vectors for each class but shared covariance matrices:

$$\Sigma_k = \Sigma \quad \forall k$$

  - Results in linear decision boundaries due to the shared covariance matrix.
- **Naive Bayes (NB):**
  - Another special case of QDA, assuming conditional independence of features given the class label:

$$p(\mathbf{x}|y = k) = \prod_{j=1}^{p} p(x_j|y = k)$$

  - This results in diagonal covariance matrices where only the variances of the features are considered, and all covariances are zero.

(b) **Assignments to Regions:** A is QDA (as it makes the most general assumptions with class-specific covariance matrices), B and C are either LDA or NB.

1) LDA can be seen as a special case of QDA if the covariance matrix is equal for all classes: $\Sigma_k = \Sigma \ \forall k$
2) Gaussian NB can be seen as a special case of QDA if the features are conditionally independent given class k:

$$p(\mathbf{x}|y = k) = p((x_1, x_2, \ldots, x_p)|y = k) = \prod_{j=1}^{p} p(x_j|y = k), \tag{1}$$

which results in diagonal covariance matrices.

3) Gaussian NB and LDA have an intersection if the covariance matrix is equal for all classes: $\Sigma_k = \Sigma \ \forall k$ **and** features are conditionally independent given class k, leaving each class with the same diagonal covariance matrix Σ.

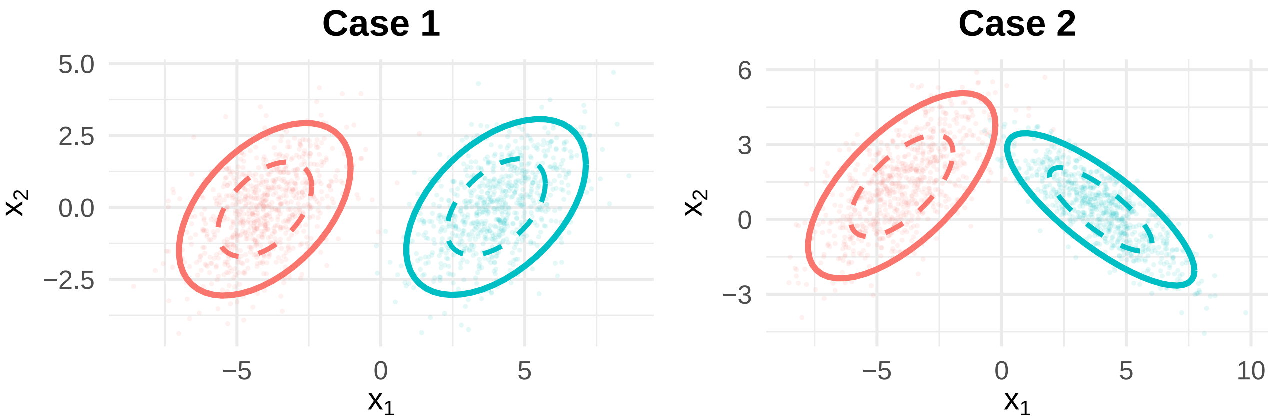
(c) **Specific Assumption:** The Venn diagram is valid under the assumption that the class-conditional distributions are Gaussian. This allows LDA, QDA, and NB to be represented as overlapping regions. In cases without Gaussian assumptions, the relationships may change, and NB may exceed the region of QDA (ellipse of NB ends outside of that of QDA to account for different distributions).

Extra task:

Below are 5 different 2D data sets illustrating different covariance structures for two Gaussian-distributed classes. The task is to assign each data set to one of the three classifiers (LDA, QDA, Gaussian NB) that can best model the data according to their assumptions. For each data set, indicate which classifier(s) can (or can not) model the data and explain why based on the covariance structure.

Case	**Class A: $\boldsymbol{\mu}_1$**	**Class B: $\boldsymbol{\mu}_2$**	**Covariance Matrices**
Case 1	$(-4, 0)$	$(4, 0)$	$\Sigma_1 = \Sigma_2 = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$
Case 2	$(-4, 1.4)$	$(4, 0.4)$	$\Sigma_1 = \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}$, $\Sigma_2 = \begin{pmatrix} 3 & -2 \\ -2 & 2 \end{pmatrix}$
Case 3	$(-4, 0)$	$(4, 1)$	$\Sigma_1 = \Sigma_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
Case 4	$(-4, 0)$	$(4, 0)$	$\Sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\Sigma_2 = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}$
Case 5	$(-4, 0)$	$(4, 0)$	$\Sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$, $\Sigma_2 = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}$

Table 1: Parameter values: means ($\boldsymbol{\mu}_k$) and covariance matrices (Σ_k) for each case.


Case 1
Case 2
x2
x1
5.0
2.5
0.0
−2.5
−5
0
5
6
3
0
−3
10

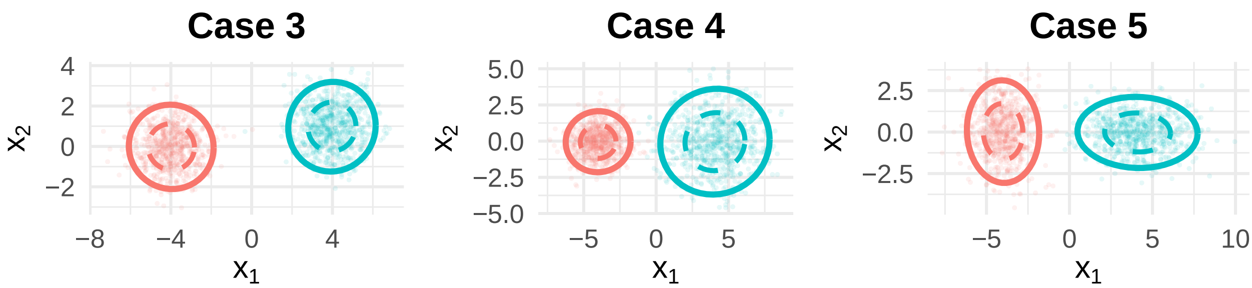
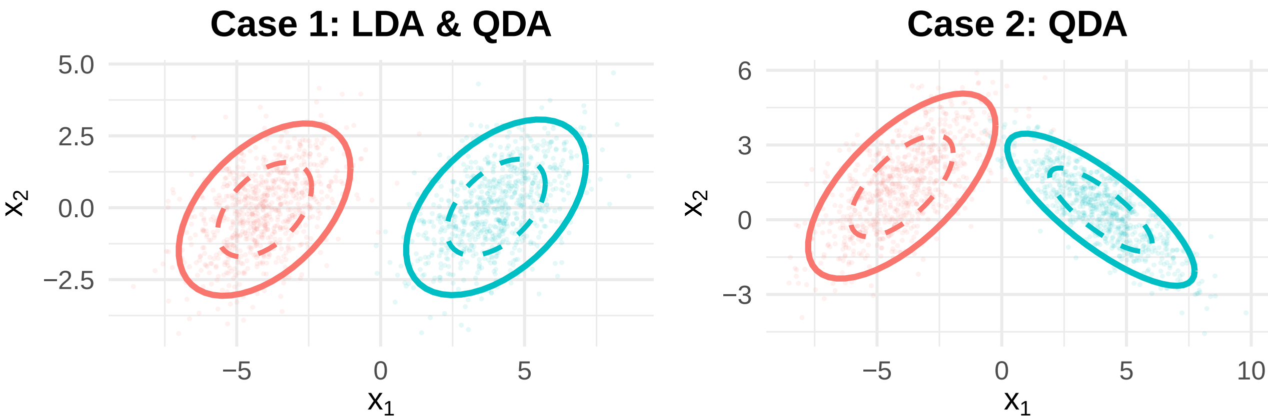

Case 3
Case 4
Case 5
x2
x1
4
2
0
−2
−8
−4
0
4
5.0
2.5
0.0
−2.5
−5.0
−5
0
5
10

Table 2: Answer key: Classification methods and covariance structures (Gaussian class-conditionals).

Case	Can_Model	Cannot_Model	Covariance_Structure	Decision_Boundary
Case 1	LDA, QDA	NB	Equal cov. matrix (equal correlation); means differ	Linear
Case 2	QDA	LDA, NB	Unequal cov. matrix (different correlation); means differ	Quadratic with cross-terms
Case 3	LDA, QDA, NB	None	Cov. matrix is diagonal and equal variances per feature (spherical); means differ	Linear
Case 4	QDA, NB	LDA	Cov. matrix is diagonal but per feature variances differ; means differ	Quadratic, axis-aligned
Case 5	QDA, NB	LDA	Cov. matrix is diagonal but only one feature has different variance; means differ	Quadratic, axis-aligned

Solution of extra task:



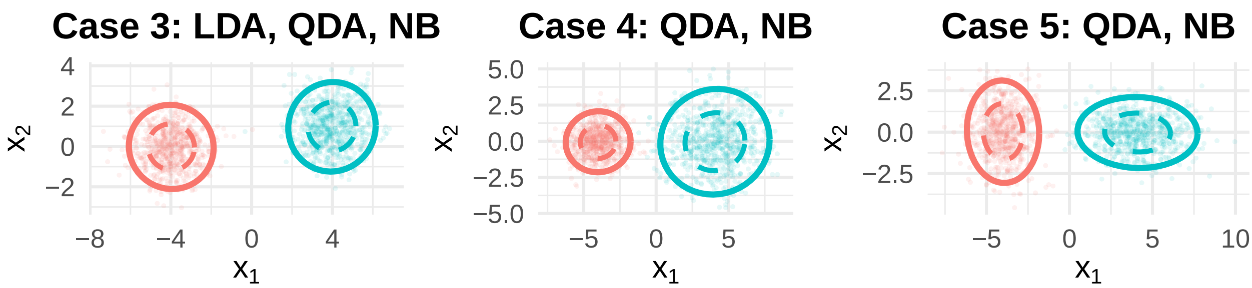


Table 3: Answer key: Explanations (Gaussian class-conditionals).

Case	Why
Case 1	Shared full covariance gives identical orientation, hence linear Bayes boundary. NB assumes diagonal covariance and cannot model correlation.
Case 2	Different full covariances create rotated ellipses, requiring cross-terms; only QDA allows class-specific full covariance.
Case 3	Spherical and equal covariances imply identical shape and scale, giving a linear boundary; NB fits since diagonal equals spherical when variances match.
Case 4	Spherical but unequal class scales yield a quadratic, axis-aligned boundary; LDA requires equal covariances and fails.
Case 5	Diagonal but unequal class covariances induce an axis-aligned quadratic boundary; NB matches the diagonal assumption; LDA requires shared covariance and fails.

Solution 2:

(a) **Linear Discriminant Analysis (LDA)**

Assuming $\mathbf{x} \mid y = k \sim \mathcal{N}(\boldsymbol{\mu}_k, \Sigma)$.

Parameters to Estimate:

- Class Priors: π_k ($K - 1$ parameters)
- Class Means: $\boldsymbol{\mu}_k$ (Kp parameters)
- Shared Covariance Matrix Σ: symmetric $p \times p$ matrix ($\frac{p(p+1)}{2}$ parameters)

Total Parameters:

$$(K-1) + Kp + \frac{p(p+1)}{2}$$

(b) **Quadratic Discriminant Analysis (QDA)**

Assuming $\mathbf{x} \mid y = k \sim \mathcal{N}(\boldsymbol{\mu}_k, \Sigma_k)$.

Parameters to Estimate:

- Class Priors: π_k ($K - 1$ parameters)
- Class Means: $\boldsymbol{\mu}_k$ (Kp parameters)
- Class Covariance Matrices Σ_k: each symmetric $p \times p$ matrix ($K \times \frac{p(p+1)}{2}$ parameters)

Total Parameters:

$$(K-1) + Kp + K \times \frac{p(p+1)}{2}$$

(c) **Naive Bayes (Gaussian Features)**

Assuming independence between features given the class:

$$x_j \mid y = k \sim \mathcal{N}(\mu_{jk}, \sigma_{jk}^2), \quad j = 1, \dots, p.$$

Parameters to Estimate:

- Class Priors: π_k ($K - 1$ parameters)
- Means: μ_{jk} (Kp parameters)
- Variances: σ_{jk}^2 (Kp parameters)

Total Parameters:

$$(K-1) + 2Kp$$