Q8. State the Bayes Decision Rule and explain its importance in classification.

Bayes Decision Rule

The Bayes decision rule is a **fundamental principle in statistical pattern recognition**. It states:

👉 *To minimize the probability of misclassification, assign an observation x to the class C_i that maximizes the posterior probability $P(C_i|x)$.*

Formally:

$$\text{Decision: } C(x) = \arg\max_i P(C_i|x)$$

Where:

- $P(C_i|x)$ = Posterior probability of class C_i given observation x.
- Derived from **Bayes' theorem**:

$$P(C_i|x) = \frac{P(x|C_i)P(C_i)}{P(x)}$$

Importance

1. **Optimality**: Minimizes classification error — no other classifier can perform better on average if the true distributions are known.
2. **Probabilistic Basis**: Provides a mathematical foundation for designing classifiers.
3. **Generalization**: Can handle multiple classes, priors, and complex likelihoods.

📌 **Example**:

- Email classification (spam vs non-spam).
- If posterior probability $P(\text{spam}|x) = 0.8$, then Bayes rule assigns the email to spam.

Q9. Suppose two classes have normal distributions with equal covariance matrices. Derive the linear discriminant function.

Setup

- Two classes: C_1, C_2.
- Features $x \in \mathbb{R}^n$.
- Class-conditional distributions:

$$p(x|C_i) \sim \mathcal{N}(\mu_i, \Sigma), \quad i = 1, 2$$

- Same covariance matrix (Σ) for both classes.

Bayesian Classifier

Posterior probability:

$$P(C_i|x) \propto p(x|C_i)P(C_i)$$

We take the **log likelihood ratio** to compare classes.

$$g_i(x) = x^T \Sigma^{-1} \mu_i - \tfrac{1}{2} \mu_i^T \Sigma^{-1} \mu_i + \ln P(C_i)$$

Linear Discriminant Function

- The decision boundary is defined by comparing $g_1(x)$ and $g_2(x)$.
- Since covariance is equal, quadratic terms cancel out → **linear boundary**.

$$g(x) = w^T x + w_0$$

Where:

- $w = \Sigma^{-1}(\mu_1 - \mu_2)$
- $w_0 = -\frac{1}{2}(\mu_1^T \Sigma^{-1} \mu_1 - \mu_2^T \Sigma^{-1} \mu_2) + \ln \frac{P(C_1)}{P(C_2)}$

📌 **Interpretation**:

- The boundary is a straight line (or hyperplane in higher dimensions).
- Intuitively, it "cuts" the feature space halfway between the means, adjusted by priors.

📌 **Example**:

- If features are student test scores, and two classes are "pass" and "fail," the discriminant function acts like a linear cutoff line (e.g., average score threshold).

Q10. Explain with an example how error probability is calculated in classification.

Error Probability

- **Error probability** is the chance that the classifier assigns a sample to the wrong class.
- In Bayes framework:

$$P_e = 1 - P(\text{correct classification})$$

Formula

For 2 classes:

$$P_e = \int \min\left[P(C_1|x), P(C_2|x)\right] p(x) dx$$

Example

Imagine medical diagnosis with two classes:

- C_1: Healthy
- C_2: Diseased

Let's say:

- $P(C_1) = 0.7, P(C_2) = 0.3$
- Classifier predicts “Healthy” if posterior probability of healthy is higher.
- If in 100 cases, 90 healthy and 10 diseased were correctly classified but 5 diseased patients were misclassified as healthy → Error probability = $5/100 = 0.05$.

📌 This shows **misclassification risk**, crucial for applications like fraud detection or cancer detection, where false negatives are very costly.

Q11. Compare linear decision boundaries and non-linear decision boundaries with diagrams/examples.

Linear Decision Boundaries

- Occur when the discriminant function is linear.
- Boundary is a **line** (2D), **plane** (3D), or **hyperplane** (nD).
- Usually arises when covariance matrices are equal (see Q9).

📌 *Example*: Logistic regression, linear SVM.

Non-Linear Decision Boundaries

- Occur when discriminant functions involve quadratic or higher-order terms.
- Emerge when covariance matrices differ, or in kernel-based methods.
- Boundaries are **curved (ellipses, circles, parabolas, etc.)**.

📌 *Example*: RBF-kernel SVM, decision trees, Gaussian mixtures.

Illustration

```
Linear Boundary (straight line):
   o o o o | x x x x
   o o o o | x x x x
             ^
             boundary (line)

Non-Linear Boundary (curve):
   o o o o o
   o     o    x
   o       x
   o o o o o x x
```

- Linear → simple, interpretable, but less flexible.
- Non-linear → flexible, but can overfit without proper regularization.

📌 **Example**:

- In handwriting recognition, linear classifiers may fail because digit shapes are non-linear. Non-linear methods (like neural nets) handle complex variations better.