Q4. A vector space is given by $V = \{(x, y) \mid x, y \in \mathbb{R}\}$. Show whether (3,4) and (6,8) belong to the same vector subspace.

Step 1: Understanding Vector Spaces

- A **vector space** is a collection of vectors where vector addition and scalar multiplication are defined and satisfy certain rules (closure, associativity, distributivity, etc.).
- Here, V represents all ordered pairs of real numbers (x, y). Geometrically, this is the entire 2D plane.

Step 2: Subspaces

- A **subspace** is a smaller “slice” of the vector space that also satisfies vector space axioms.
- Example: A line through the origin in $\mathbb{R}^2$ is a subspace.

Step 3: Relation Between (3,4) and (6,8)

- To check if they belong to the same subspace, we ask:

  > Is one vector a scalar multiple of the other?

- Compute:

$$(6, 8) = 2 \times (3, 4)$$

- Yes, (6,8) is exactly double (3,4).

Conclusion

- Both lie on the line defined by all multiples of (3,4).
- That line (passing through the origin and the point (3,4)) is a **1-dimensional subspace** of $\mathbb{R}^2$.

📌 **Example in Pattern Recognition**: In face recognition, feature vectors often lie in a lower-dimensional subspace (like eigenfaces in PCA). Two images of the same person under different lighting may be scalar multiples of each other, thus lying in the same subspace.

Q5. What are the basic axioms of probability? Give one example from pattern recognition.

Kolmogorov’s Axioms of Probability

1. **Non-negativity**: For any event A,

$$P(A) \geq 0$$

   Probability can never be negative.

2. **Normalization**: For the sample space S,

$$P(S) = 1$$

   The probability of all possible outcomes is always 1.

3. **Additivity**: For any two disjoint events A and B,

$$P(A \cup B) = P(A) + P(B)$$

If two events cannot occur together, their combined probability is the sum of individual probabilities.

Example in Pattern Recognition

- Suppose a classifier is predicting whether an email is **spam (S)** or **not spam (N)**.
- Sample space $S = \{S, N\}$.
- The model outputs:
  - $P(S) = 0.7$,
  - $P(N) = 0.3$.
- Here:
  - Non-negativity: both values ≥ 0 ✅
  - Normalization: $P(S) + P(N) = 1$ ✅
  - If events were disjoint categories (spam vs not spam), additivity holds ✅

📌 This ensures our probabilistic model behaves consistently with mathematical foundations.

Q6. Explain the concept of estimation theory and its role in pattern recognition.

What is Estimation Theory?

- Estimation theory deals with methods to estimate **unknown parameters** of a distribution or model based on observed data.
- In statistics and ML, we rarely know the true distribution of data; instead, we estimate its parameters.

Two Common Estimators

1. **Maximum Likelihood Estimation (MLE)**:
   - Finds parameter values that maximize the likelihood of observed data.
   - Example: Estimating mean and variance of a Gaussian distribution for class features.
2. **Bayesian Estimation**:
   - Incorporates prior knowledge with observed data.
   - Produces a *posterior distribution* of parameters.

Role in Pattern Recognition

- Classifiers (like Naïve Bayes or HMMs) need probability distributions. These distributions have parameters (means, variances, transition probabilities).
- Estimation theory helps **learn these parameters** from training data.

📌 **Example**:

- In digit recognition, suppose feature = “height of digit bounding box.”
- If we assume it follows a normal distribution:
  - Use training data of digit "5" images to estimate mean μ and variance σ^2.
  - Classifier then uses these estimated values for decision-making.

Without estimation theory, we cannot train probabilistic models effectively.

Q7. Discuss the role of metric spaces and distances in defining decision regions. Illustrate using Euclidean and Mahalanobis distance.

Decision Regions in Pattern Recognition

- A **decision region** is the portion of the feature space where all points are assigned to the same class.
- Boundaries between regions are defined using **distance measures** (metrics).

Metric Spaces

- A metric space is a set with a distance function $d(x, y)$ satisfying:
  1. Non-negativity
  2. Identity of indiscernibles ($d(x, y) = 0 \iff x = y$)
  3. Symmetry ($d(x, y) = d(y, x)$)
  4. Triangle inequality ($d(x, z) \leq d(x, y) + d(y, z)$)

Distance Measures

1. **Euclidean Distance (L2 norm)**

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

   - Measures straight-line distance in feature space.
   - Decision boundary: perpendicular bisector between two class means.

📌 *Example*:

- If two fruit classes (apple vs orange) are represented by features (weight, diameter), Euclidean distance assigns a new fruit to the nearest centroid.

---

2. **Mahalanobis Distance**

$$d(x, \mu) = \sqrt{(x - \mu)^T \Sigma^{-1} (x - \mu)}$$

   - Considers correlation between features via covariance matrix Σ.

   - Useful when feature dimensions have different scales or correlations.

📌 *Example*:

- In face recognition, pixel intensities are correlated. Mahalanobis distance accounts for this correlation, unlike Euclidean.

Illustration of Decision Regions

- **Euclidean** → Circular decision boundaries around class means.
- **Mahalanobis** → Elliptical boundaries that adapt to feature correlations.

```
Euclidean:       Mahalanobis:
   ○                ⬭
  ○ ○              ⬭⬭
   ○                ⬭
```

Conclusion

- Distances define how classifiers separate classes.
- Euclidean is simple but assumes independence and equal variance.
- Mahalanobis is more powerful for correlated, real-world data.

📌 *Example*: In medical diagnosis, Mahalanobis distance can better separate patient groups when symptoms/features are correlated.