Lecture Summary: Covariance and Correlation

Source: Lecture 3.5.pdf

Key Points

- **Motivation:**
  - Variance measures the spread of a single random variable, while covariance and correlation quantify the relationship between two random variables.
  - Example:
    - Two joint PMFs can have identical marginals but very different relationships between the random variables.
    - Covariance and correlation help capture these relationships.
- **Covariance:**
  - Definition:
    $$\text{Cov}(X, Y) = E\left[(X - E[X])(Y - E[Y])\right].$$
  - Simplified formula:
    $$\text{Cov}(X, Y) = E[XY] - E[X]E[Y].$$
  - Interpretation:
    - $\text{Cov}(X, Y) > 0$: X and Y tend to increase together.
    - $\text{Cov}(X, Y) < 0$: When X increases, Y tends to decrease.
    - $\text{Cov}(X, Y) = 0$: X and Y are uncorrelated.
- **Examples of Covariance:**
  - **Positive Covariance:**
    - X: Height of a person, Y: Weight of a person.
    - Taller individuals tend to weigh more.
  - **Negative Covariance:**
    - X: Rainfall during monsoon, Y: Farmer debt.
    - Higher rainfall correlates with lower farmer debt.
- **Correlation:**
  - Normalized version of covariance:
    $$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)},$$
    where $\sigma(X)$ and $\sigma(Y)$ are the standard deviations of X and Y.
  - Properties:
    - $-1 \leq \rho(X, Y) \leq 1$.

    * $\rho(X, Y) = 1$: Perfect positive correlation.
    * $\rho(X, Y) = -1$: Perfect negative correlation.
    * $\rho(X, Y) = 0$: No linear relationship.

- **Properties of Covariance:**
  1. $\text{Cov}(X, X) = \text{Var}(X)$.
  2. $\text{Cov}(X + a, Y) = \text{Cov}(X, Y)$ (translation invariance).
  3. $\text{Cov}(aX, bY) = ab \cdot \text{Cov}(X, Y)$ (scaling).
  4. Symmetry:
  $$\text{Cov}(X, Y) = \text{Cov}(Y, X).$$

- **Relation Between Covariance and Independence:**
  - Independence implies $\text{Cov}(X, Y) = 0$ (uncorrelated).
  - However, $\text{Cov}(X, Y) = 0$ does not imply independence.
  - Example of dependent but uncorrelated variables:
    * $X \in \{-1, 0, 1\}$, $Y \in \{0, 1\}$ with specific joint probabilities.
    * $E[XY] = E[X]E[Y]$, but X and Y are not independent.

Simplified Explanation

Covariance: Measures how two random variables move together: - Positive: Variables increase together. - Negative: One increases while the other decreases.

Correlation: A normalized measure of covariance indicating the strength of a linear relationship.
Examples: - Positive covariance: Height and weight. - Negative covariance: Rainfall and farmer debt.
Important Note: Uncorrelated variables may still be dependent.

Conclusion

In this lecture, we:

- Defined covariance and correlation as measures of relationships between random variables.
- Explored their properties and practical examples.
- Highlighted the difference between independence and uncorrelation.

Covariance and correlation are foundational tools for understanding relationships in multivariable data.