Lecture 8: Information Theory in Pattern Recognition

Dr. Pooja Sarin

21/22 January, 2025

1 Introduction to Information Theory

Information theory is a mathematical framework for quantifying the transmission, processing, and storage of information. It plays a critical role in pattern recognition by helping quantify uncertainty, relevance, and dependencies in data.

2 Basics of Information Theory

2.1 Entropy

Entropy measures the uncertainty or randomness in a system. For a discrete random variable X with probability distribution $P(x)$:

$$H(X) = -\sum_{x \in X} P(x) \log P(x) \tag{1}$$

- **High Entropy:** Indicates greater uncertainty (e.g., a fair coin toss).
- **Low Entropy:** Indicates less uncertainty (e.g., biased outcomes).

Example: For a fair coin, $P(Heads) = P(Tails) = 0.5$, so:

$$H(X) = -[0.5 \log 0.5 + 0.5 \log 0.5] = 1\,\text{bit} \tag{2}$$

2.2 Mutual Information

Mutual Information quantifies the amount of information one random variable contains about another. For random variables X and Y:

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} P(x,y) \log \frac{P(x,y)}{P(x)P(y)} \tag{3}$$

- **High Mutual Information:** Indicates strong dependency between X and Y.
- **Low Mutual Information:** Indicates weak or no dependency.

Example: In a dataset of customer purchases, mutual information can quantify how strongly a customer's age predicts their buying behavior.

3 Applications in Pattern Recognition

3.1 Feature Selection

Information theory is widely used in feature selection to identify the most relevant features for a predictive model.

- **Entropy-Based Methods:** Features with high information gain (reduction in entropy) are prioritized.
- **Mutual Information-Based Methods:** Features with high mutual information with the target variable are selected.

Example: In text classification, mutual information can rank words based on their relevance to class labels, helping reduce the dimensionality of the feature space.

3.2 Clustering

Information theory supports clustering by quantifying the similarity or dissimilarity between clusters.

- **Entropy in Clustering:** Measures the homogeneity of clusters. Lower entropy indicates better-defined clusters.
- **Mutual Information in Clustering:** Measures the overlap between predicted and actual clusters (e.g., Adjusted Mutual Information).

Example: In image segmentation, mutual information can compare how well pixel clusters align with ground truth labels.

4 Illustrations and Examples

4.1 Feature Selection Example

Dataset: Binary classification problem with two features, X_1 (relevant) and X_2 (irrelevant).

- Calculate the mutual information $I(X_1; Y)$ and $I(X_2; Y)$ with the target Y.
- Select X_1 if $I(X_1; Y) > I(X_2; Y)$.

Outcome: The irrelevant feature X_2 is excluded, simplifying the model.

4.2 Clustering Example

Task: Segment customers into groups based on purchase patterns.

- Use entropy to evaluate the homogeneity of clusters.
- Use mutual information to compare clustering results with predefined categories (e.g., VIP vs. regular customers).

Outcome: A clustering algorithm with lower entropy and higher mutual information is preferred.

5 Conclusion

Information theory provides essential tools for quantifying uncertainty, relevance, and dependencies in data. By leveraging concepts like entropy and mutual information, it supports critical tasks in pattern recognition such as feature selection and clustering. These techniques enable more efficient and interpretable models, enhancing the overall performance of machine learning systems.