Nonparametric Techniques

- Approximate a “nice” function locally by simpler functions
- Given iid samples \( \{X_1, \ldots, X_n\} \), \( X_i \sim p(x) \)
- Basic idea: counting

Simple Example

\( X \sim N(0,1) \)

Density Estimation

- The probability a vector falling in a region
  \[ P = \int p(x) dx \]
- Given n iid samples, \( K \): # of samples in \( R \)
  \[ P(K = k) = \binom{n}{k} P^k (1 - P)^{n-k} \]
  \[ E[K] = nP \]
Density Estimation

- Assume continuous density function and small $R$

$$\int_{\mathbb{R}^d} p(x)dx = p(x)V$$

$$p(x) = \frac{k}{V}$$

Density Estimation

- A theoretical standpoint – unlimited samples

$$p_n(x) = \frac{n}{V_n}$$

- For convergence, it requires

$$\lim_{n \to \infty} p_n = 0$$

$$\lim_{n \to \infty} k_n = \infty$$

$$\lim_{n \to \infty} \frac{k_n}{p_n} = 0$$

Parzen Windows

- Consider a hypercube $R_n \subset \mathbb{R}^d$ with volume $V_n = h_n^d$

- Define a window function

$$\varphi(u) = \begin{cases} 1 & |u| \leq 1/2 \\ 0 & \text{otherwise} \end{cases}$$
Parzen Windows

- The number of samples in a hypercube
  \[ k_i = \sum_{j=1}^{n} \phi \left( \frac{x_i - x_j}{h} \right) \]
- An estimate
  \[ p_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{V_x} \phi \left( \frac{x - x_i}{h} \right) \]

K-Nearest Neighbor Estimation

- Fix \( k_n \)

Discussion

- Generality
- Computation time
- Storage
- Curse of dimensionality
The Nearest-Neighbor Rule

• A test point belongs to the class of its nearest neighbor
• Convergence of the nearest neighbor rule

\[ P' \leq P \leq P' \left(2 - \frac{c}{\epsilon - 1}\right) \]

The Nearest-Neighbor Rule

• Probability of error

\[ p(e) = P(e | x) = \int p(e | x) p(x) dx \]

• Bayesian error

\[ p^*(e | x) = 1 - P(\omega_k | x) \]

\[ m = \arg \max \ p(\omega_k | x) \]

\[ P^*(e) = \int p^*(e | x) p(x) dx \]

The Nearest-Neighbor Rule

• Conditional error of the NN-rule

\[ p_{\omega_k}(e | x) = \sum_{\omega_k} p(\omega_k | x) [1 - p(\omega_k | x)] \]

• Error of the NN-rule

\[ E_{\omega_k}(e) = \int p_{\omega_k}(e | x) p(x) dx \]
The Nearest-Neighbor Rule