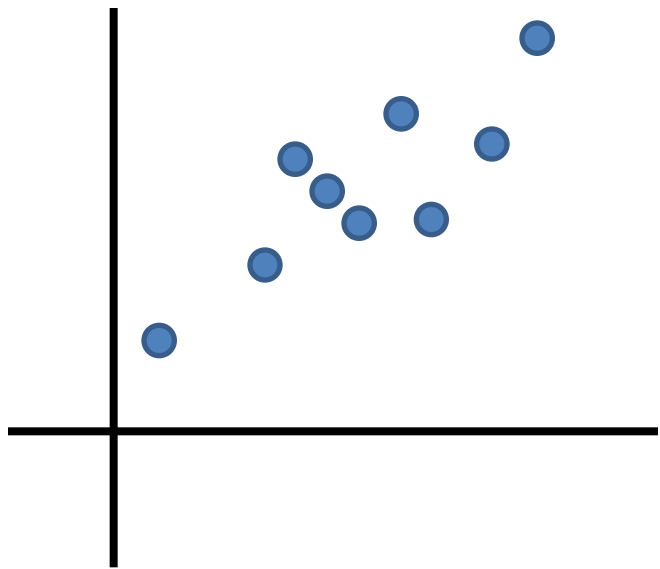


「パターン認識と機械学習」

2-5 ノンパラメトリック法

新納浩幸

パラメトリックとノンパラメトリック



$$y = f(x) = ?$$

パラメトリック

$$f(x) = ax + b$$

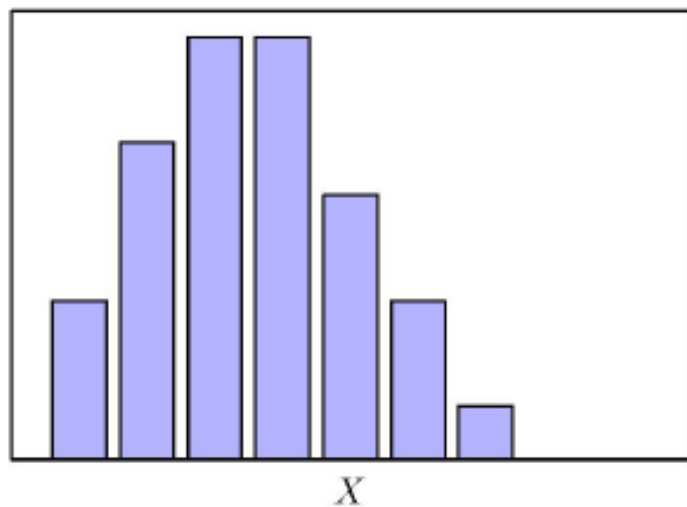
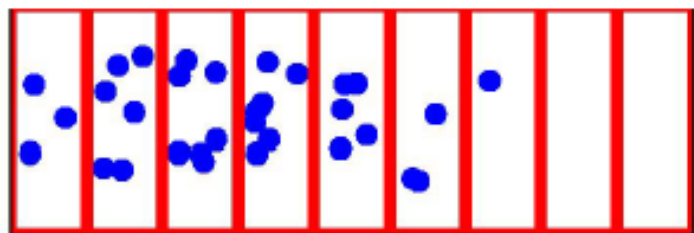
パラメータは a と b

ノンパラメトリック

$$f(x) = ??$$

データに依存して決定する
(パラメータが無限個)

ヒストグラム密度推定法



x : 連続変数

Δ_i : i 番目の幅

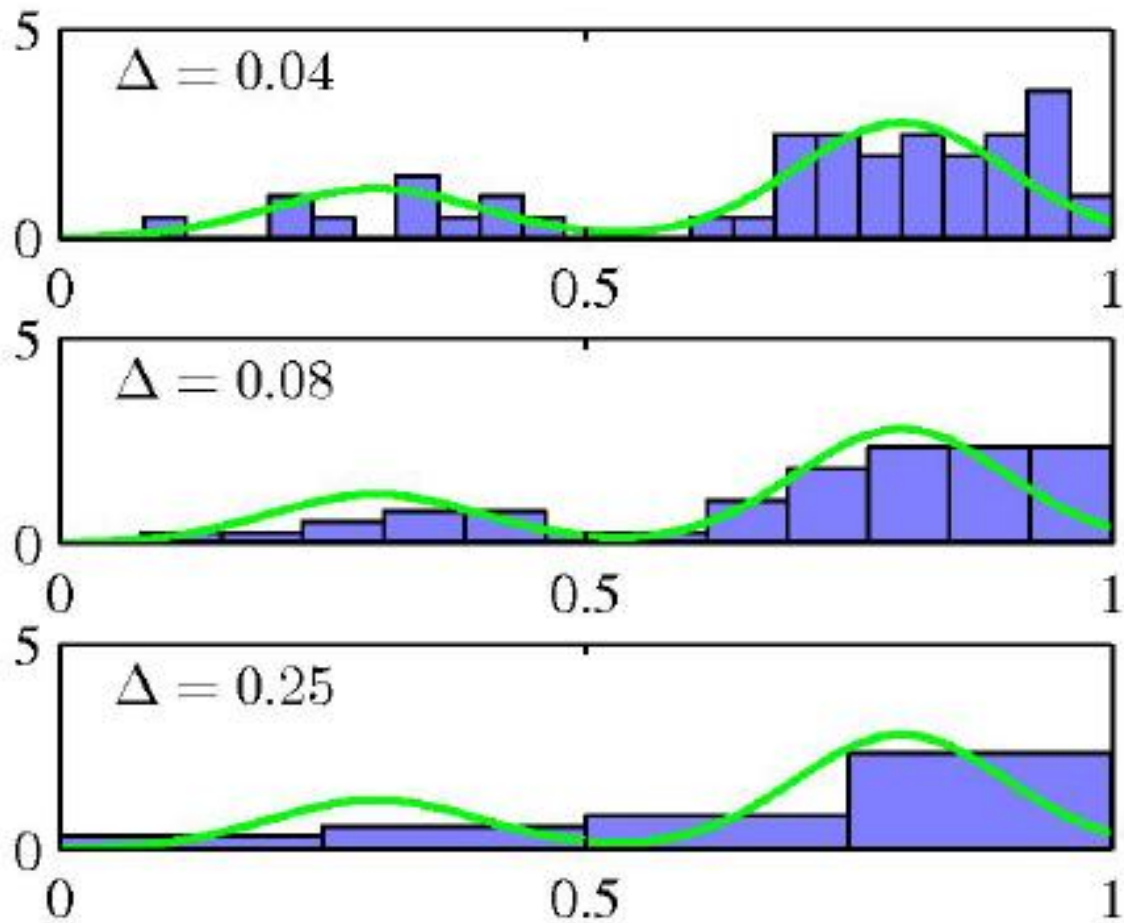
n_i : i 番目の観測値の数

N : 観測値の総数

$$\text{確率密度: } p_i = \frac{n_i}{N\Delta_i}$$

Δ の影響

Δ は適切に設定する必要あり



ヒストグラム密度推定法の長短

利点	欠点
<ol style="list-style-type: none">一度ヒストグラムを求めると元データを破棄できる。 (大規模なデータに有利)データが逐次的に与えられた時に容易に適用できる。	<ol style="list-style-type: none">推定した密度が区間の縁で不連続になる。次元数が増えると、指数的に区間の総数が増え、計算規模が増加する。

実際には、1次元か2次元のデータの簡単な可視化には役立つが、他のほとんどの密度推定の応用問題には適してない。

ノンパラメトリックのまとめ

1. 特定の位置の確率密度を推定するにはその点の近傍のデータ点も考慮する。
 - 近傍の特性は区間によって定義されている。
(区間=ヒストグラムアプローチの平滑化パラメータ)
2. 平滑化パラメータの値は大きすぎも小さすぎもしない、適切な値でないといけない。