$Column \sim$ ノンパラメトリックモデル（カーネル密度推定と最近傍法）$\sim$

$$P = \frac{K}{N}$$

$$P \approx Vp(x)$$

よって上式より

$$p(x) \approx \frac{K}{NV}$$

ここで２つのアプローチが考えられる。

1.Vを先に決めて K を数える。（カーネル密度推定）

2.K を与えて、Vを決定する。（最近傍法、K 近傍法）

1. カーネル密度推定

$$V = h^D$$

$$W(u) = \begin{cases} 1 & |u_i| \leq 1/2 \\ 0 & else \end{cases}$$

$$K = \sum_{i=1}^{n} W\Big[\frac{x - x_i}{h}\Big]$$

よって推定密度は以下のようになる。

$$p(x) = \frac{1}{N}\sum_{n=1}^{N} \frac{1}{h^D} W\Big[\frac{x - x_i}{h}\Big]$$

一般的にはこのように立方体の中にデータ点をいれるのではなく、中心がデータ点 x_n にある N 個の立方体を重ね合わせるという解釈する。

このカーネルは立方体の縁で不連続になるので、ガウスカーネルを用いて以下のような確率密度モデルを考えることが多い。（中心がデータ点 x_n にある N 個のガウス分布を重ね合わせる。）

$$p(x) = \frac{1}{N}\sum_{n=1}^{N} \frac{1}{(2\pi h^2)^{D/2}} \exp\Big(-\frac{||x - x_n||^2}{2h^2}\Big)$$

2. **最近傍法**

今度は先にデータ数 K を決めて、それに応じて Vを決定する。ここでは立方体ではなく d 次元の球（超球）を考える。

超球の体積は

$$V = \frac{\pi^{\frac{d}{2}} r^d}{\Gamma(\frac{d}{2}+1)}$$

よって（）より確率密度は、

$$p(x) = \frac{K\Gamma(\frac{d}{2}+1)}{N_k \pi^{\frac{d}{2}} r^d}$$

ここで $K = 1$, 最も近いデータ点までの距離を Δ とすると、

$$p(x|C_k) = \frac{\Gamma(\frac{d}{2}+1)}{N\pi^{\frac{d}{2}}\Delta^d}$$

これにベイスの定理を適用してクラスの事後分布を求めると、

$$p(C_k|x) = \frac{p(x|C_k)p(C_k)}{p(x)} \propto p(x|C_k)p(C_k)$$

$$\approx \frac{\Gamma(\frac{d}{2}+1)}{N_k\pi^{\frac{d}{2}}\Delta^d}\frac{N_k}{N} \propto \frac{1}{\Delta^d}$$

よって、データ点からの距離が近いほどクラスの事後分布が大きくなる。このことから、データ点から一番近いデータのクラスを選択すればいいことがわかる。よってこの手法を最近傍法と呼ぶ。これを一般の K に拡張したものが K 近傍法である。（PRML 上巻第 2 章参照）