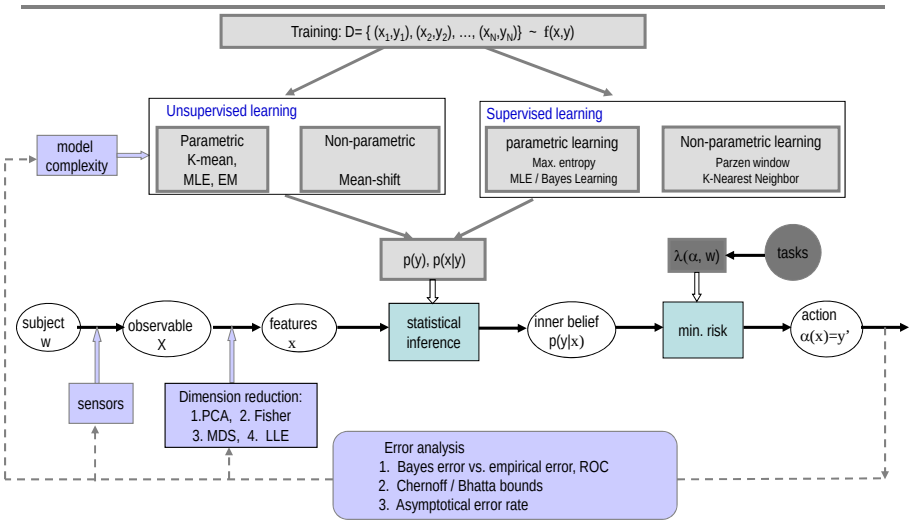
## Diagram of Bayesian Method for Pattern Classification






## Lecture 17: Non-parametric learning with Parzen Windows

1. Parzen window.

2. Convergence analysis

Motivation:

When we have a large number of object categories, say 1000.
It becomes hard to train classifiers like Boosting/SVM, instead, non-parametric methods, such as nearest neighbor classifiers are more effective.

For example, handwriting Chinese character recognition on hand hold devices.

# Density / model estimation

In practice, many class models are so complicated that it is extremely hard to develop parametric Forms. For example,

Estimate the distribution of annual rainfall / meteoric stones in the United States. Ignoring seasonal variation, p(x,y) is the probability that a raindrop hits a position (x,y).

Difficulties:

1. multi-modal density --- p(x,y) will have many peaks, thus it is hard for parametric models.
2. it is difficult to collect a large reasonable amount of data at each point.



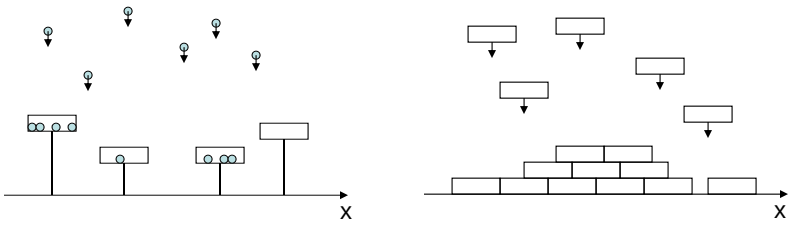
# Intuitive ideas

Given: a set of training data which are iid samples from an underlying frequency for each class

$D=\{ x_1, x_2, \ldots, x_n \}$

Goal: to estimate the class density / model p(x).

Two views: one is centered at the point x, and one is centered of the training examples.



## Non-parametric density estimation

Suppose we take the view centered at each point x.

For each point x, we form a window centered at x, whose volume is $V_n$. We count the number of samples $k_n$ from D that fall in the window.

The probability of falling in the box is $P_n = \frac{k_n}{n}$

The density is estimated as

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

The subscript n indicates that this estimation depends on the size of the training set n.



## Non-parametric density estimation

Goal: to design a sequence of windows $V_1$, $V_2$, $V_n$, .... at point x, so that

$$as \;\; n \to \infty, \;\; p_n(x) \to f(x)$$

Conditions for the window design

1. $\lim_{n\to\infty} V_n = 0,$ increasing spatial resolution
2. $\lim_{n\to\infty} k_n = \infty,$ assuming $p(x) \neq 0$ and large samples at each point.
3. $\lim_{n\to\infty} k_n / n = 0,$ $k_n$ grows in a order smaller than n.

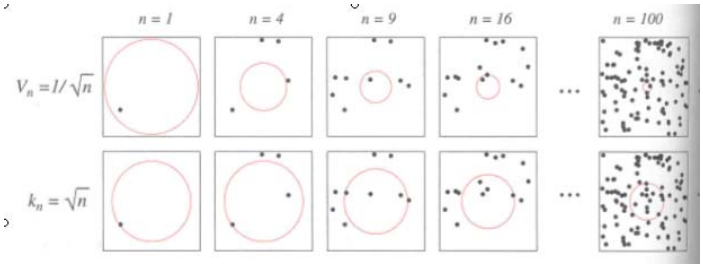
## Two general methods for window design

1. Parzen window: fixing window size for all positions x.

$$V_n = 1/\sqrt{n}$$

2. k-NN: Adapting window locally according to density, so more precision for denser places

$$k_n = \sqrt{n}$$






## Parzen Window

Parzen window is a window function

$$\varphi(u) \geq 0, \quad \int_{\Omega} \varphi(u)\, du = 1$$

Example 1: an unit hypercube in $R^d$

$$\varphi(u) = 1, \text{ if } |u| \leq 1/2, \quad = 0 \text{ otherwise}$$

Example 2: a Gaussian function

$$\varphi(u) = \frac{1}{(2\pi)^d} e^{-u^2/2}$$

The window can be scaled, translated, and normalized, let V be the effective volume of the window.

$$\delta(u - x_i) = \frac{1}{V} \varphi\left(\frac{u - x_i}{h}\right)$$

# Parzen window

For parzen window, we can estimate the density in the following way:

Choosing a scale $h_n$

$$k_n(x) = \sum_{i=1}^{n} \varphi(\frac{x - x_i}{h_n})$$

$$p_n(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{V_n}\varphi(\frac{x - x_i}{h_n}) = \frac{1}{n}\sum_{i=1}^{n} \delta(x - x_i)$$



# Example

Estimating a density with 5 samples, or Gaussian windows of various scales h=1, 0.5, 0.2

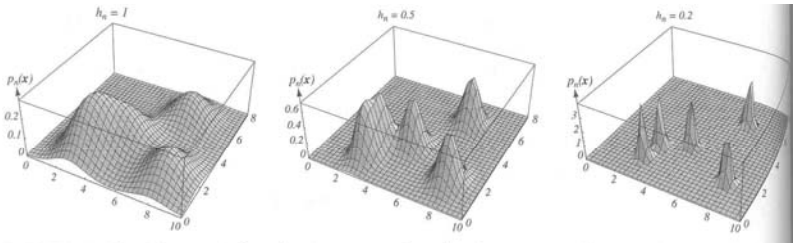
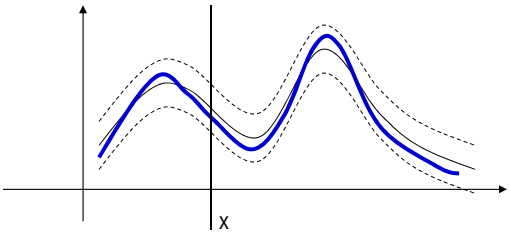


FIGURE 4.4. Three Parzen-window density estimates based on the same set of five samples, using the windo functions in Fig. 4.3. As before, the vertical axes have been scaled to show the structure of each distribution

## Convergence proof

We can show that the Parzen window estimation converge to the true density at every point x

$$p_n(x) = \frac{1}{n}\sum_{i=1}^{n} \delta(x - x_i) \;\rightarrow\; f(x) \quad or \quad p(x), \qquad as \;\; n \rightarrow \infty$$




At each point x, we prove the convergence of mean and the variance decreases to zero



## Convergence of mean

The expectation is w.r.t. the underlying distribution of the samples.

$$f(x_1, \ldots x_n) = f(x_1) \cdots f(x_n)$$

$$\begin{aligned} E_{f(x_1,\ldots x_n)}[p_n(x)] &= \frac{1}{n}\sum_{i=1}^{n} E_{f(x_1,\ldots x_n)}[\delta_n(x - x_i)] \\ &= \frac{1}{n}\sum_{i=1}^{n} E_{f(x_i)}[\delta_n(x - x_i)] \\ &= E_{f(y)}[\delta_n(x - y)] \\ &= \int f(y)\delta_n(x - y)\,dy \end{aligned}$$

Clearly, the expectation of the estimation converges to a convolution of the true density with the Parzen window. As n increases to infinity. The window becomes a Dirac delta function. Then the expectation is f(x).

## Convergence of variance

$$\sigma_n^2(x) = E_{f(x_1,\ldots x_n)}[(p_n(x) - E_f[p_n(x)])^2]$$
$$= \frac{1}{n}(E_f[\delta_n^2(x-y)] - E_f^2[p_n(x)])$$
$$= \frac{1}{nV_n}\int \frac{1}{V_n}\varphi^2(\frac{x-y}{h_n}) f(y)dy - \frac{1}{n}E_f^2[p_n(x)]$$
$$\leq \frac{\sup(\varphi(u))E[p_n(x)]}{nV_n} \to 0$$



## Example:

The underlying density is a Gaussian. The window volume decreases as n increases.

$$V_n = V_1/\sqrt{n}$$

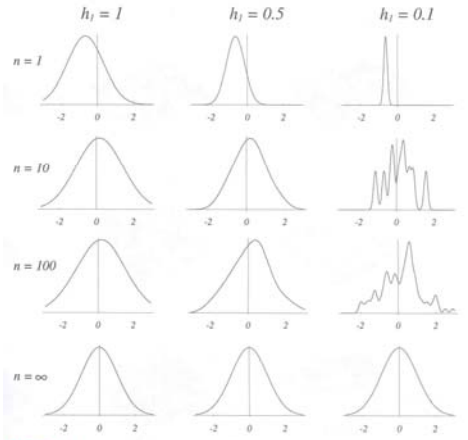



FIGURE 4.5. Parzen-window estimates of a univariate normal density using different window widths and numbers of samples. The vertical axes have been scaled to best show the structure in each graph. Note particularly that the $n = \infty$ estimates are the same (and match the true density function), regardless of window width.

# Example

The underlying density is a two-modal. Although h1 is different, they converge to the same true density

$$V_n = V_1 / \sqrt{n}$$

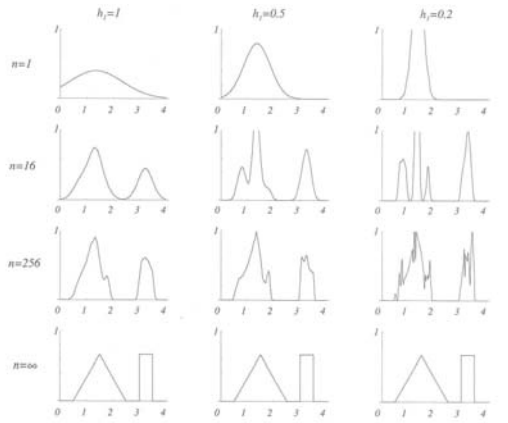



FIGURE 4.7. Parzen-window estimates of a bimodal distribution using different window widths and numbers of samples. Note particularly that the $n = \infty$ estimates are the same (and match the true distribution), regardless of window width.