

شناسایی الگو
فصل دوم
روش‌های بدون پارامتر
NONPARAMETRIC MODELS

Textbook: Pattern Recognition, S. Theodoridis , K. Koutroumbas, Fourth Edition, 2009.

مدل‌های پارامتری

○ اگر $P(\omega_i)$ و $p(x|\omega_i)$ معلوم باشند، روش بیز، طبقه‌بند بهینه به دست می‌دهد.

○ اگر نوع توزیع معلوم باشد، اما پارامترهای آن نامعلوم باشد، با داشتن تعدادی داده آموزش و با استفاده از یکی از دو روش

- Maximum-Likelihood Estimation (MLE)
- Maximum a Posteriori probability estimation (MAP)

می‌توان پارامترها را تخمین زد. خروجی این دو روش تقریباً یکی است.

مدل‌های بدون پارامتری

○ تخمین چگالی با مدل‌های پارامتری فرض می‌کند که نوع توابع چگالی معین است. با این حال، فرم‌های پارامتری رایج همیشه با چگالی‌هایی که در عمل با آن مواجه می‌شوند، مطابقت ندارند. روش‌های ناپارامتریک را می‌توان برای توزیع‌های دلخواه استفاده کرد.

مدل‌های بدون پارامتری – روش هیستوگرام

○ یک روش بسیار ساده این است که فضا به تعدادی سلول (bins) با اندازه مساوی تقسیم شود و هیستوگرام محاسبه شود.

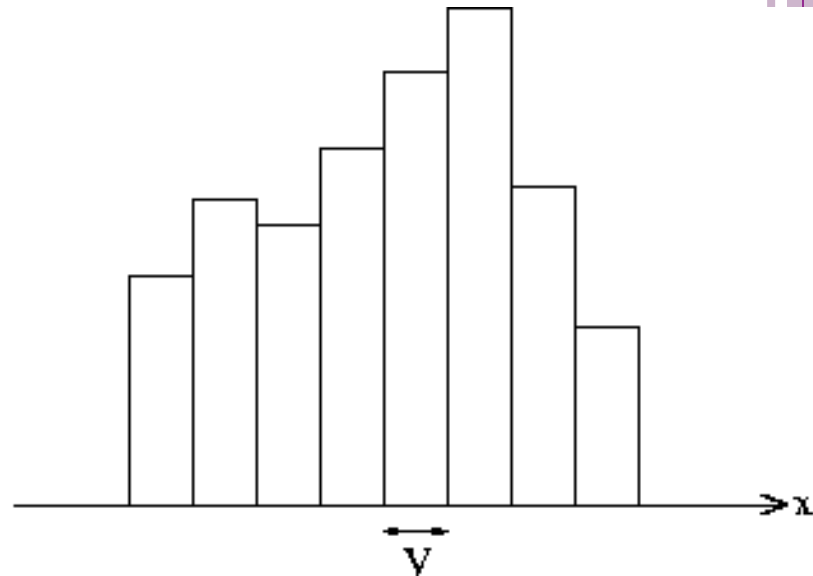
○ تابع چگالی در نقطه x این گونه محاسبه می‌شود:

$$P_n(x) = \frac{k}{nV}$$

n = number of samples.

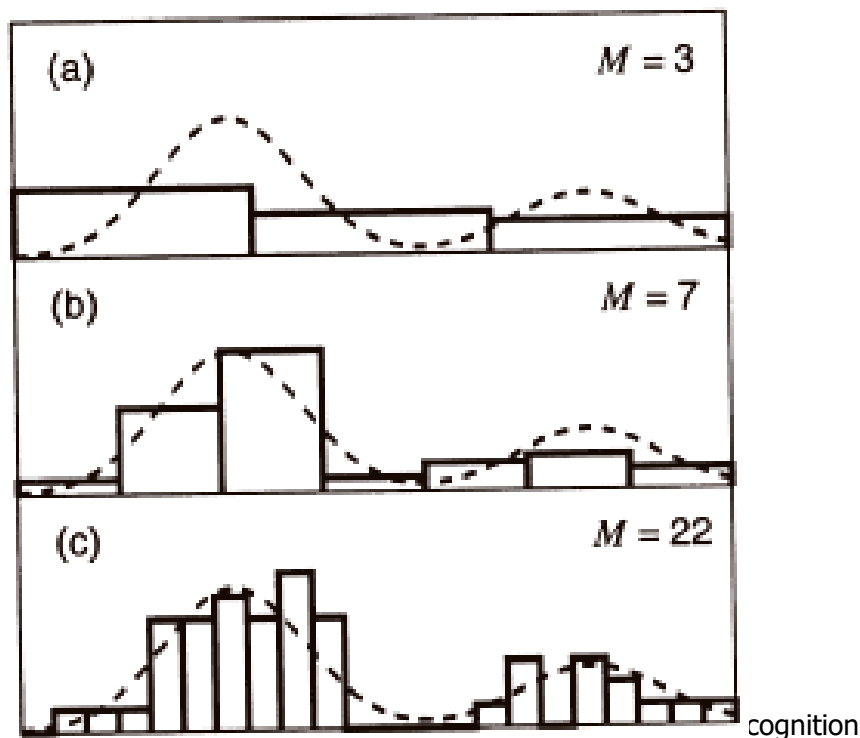
k = number of samples
in the cell that includes x .

V = is the volume of that cell.



مدل‌های بدون پارامتری - روش هیستوگرام - محاسبه تابع چگالی

- تعداد سلول‌ها که با M نمایش داده می‌شود؛ به عنوان یک پارامتر هموارسازی عمل می‌کند.
- اگر عرض سلول کوچک باشد؛ چگالی تخمین زده شده بسیار تیز خواهد بود.
- اگر عرض سلول بزرگ باشد؛ ساختار واقعی چگالی هموار می‌شود.
- در عمل، باید یک مقدار بهینه برای M پیدا شود تا تعادل ایجاد نماید.



مدل‌های بدون پارامتری – روش هیستوگرام – مزایا و معایب

مزایا

- هنگامی که هیستوگرام به دست آمد، دیگر به داده‌ها نیازی نیست (حافظه کارآمد)
- با یکبار دیدن داده‌ها می‌توان هیستوگرام را ساخت.

معایب

- چگالی به دست آمده هموار نیست و دارای ناپیوستگی در مرزهای سلول‌های هیستوگرام است.
- اگرچه پیاده‌سازی بسیار آسانی دارد، اما به دلیل تعداد سلول‌ها معمولاً در فضاهای با ابعاد بالا عملی نیست.
- مشاهدات زیادی برای جلوگیری از صفر بودن تخمین در یک منطقه بزرگ مورد نیاز است.

مدل‌های بدون پارامتری – روش هیستوگرام – تخمین پارامترها

$$P_n(x) = \frac{k}{nV}$$

چگونه مقدار مناسب برای دو پارامتر k و V انتخاب شود؟

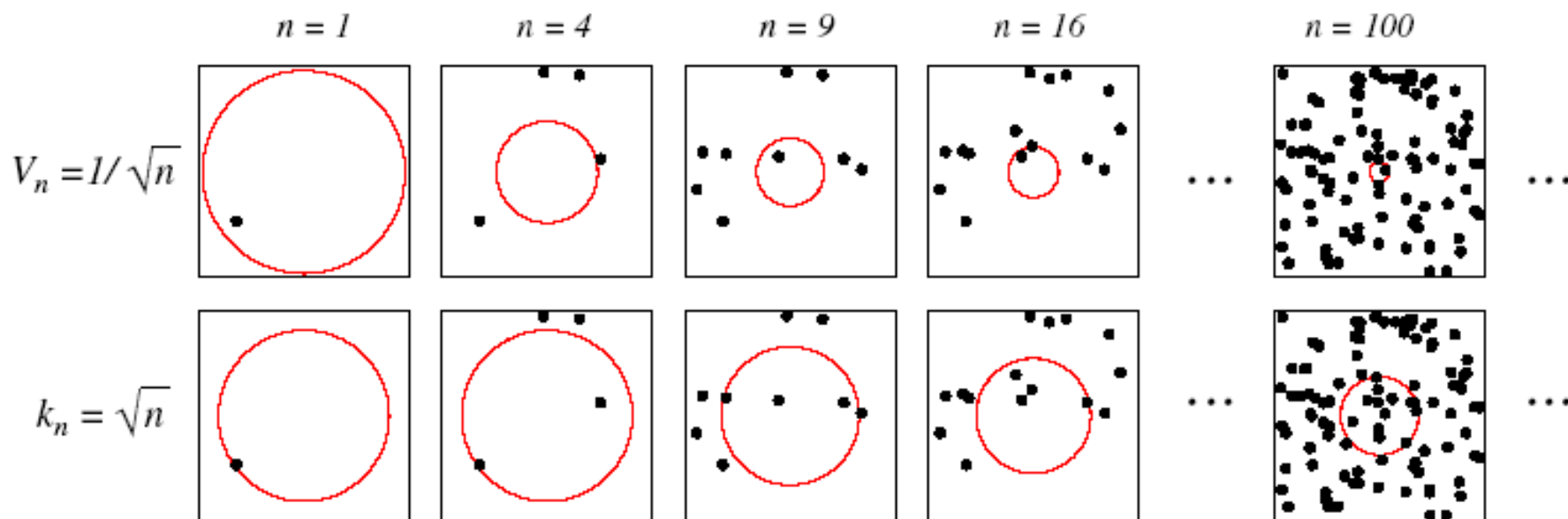
روش اول: اندازه ناحیه تابعی از تعداد نمونه باشد (Parzen Window).

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

روش دوم: مقدار k_n تابعی از مقدار نمونه باشد (K-Nearest Neighbor).

$$k_n = \sqrt{n}$$

مدلهای بدون پارامتری - روش هیستوگرام - تخمین پارامترها

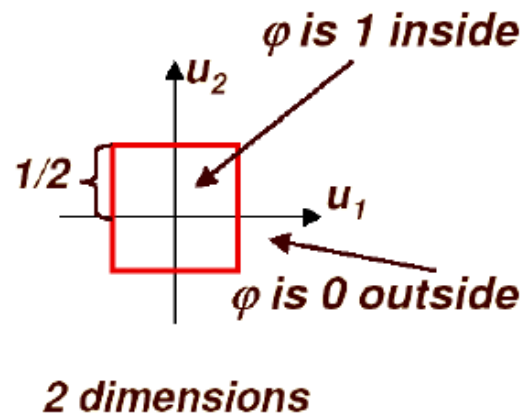
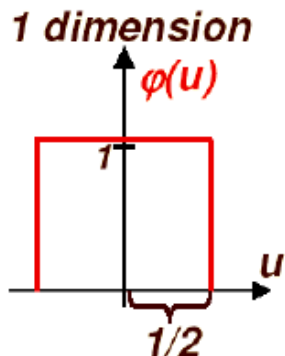


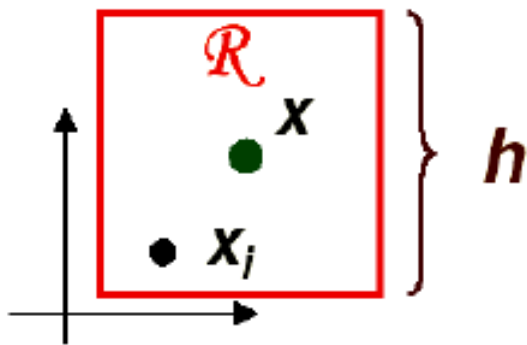
روش‌هایی برای تخمین چگالی و کلاس‌بندی مرکز مربع.

پنجره پارزن

○ برای یافتن تعداد نمونه‌ها در یک ناحیه، یک تابع هسته (کرنل) تعریف می‌شود. این کرنل که یک ابر مکعب به حجم واحد در مرکز مبدا است، به عنوان پنجره پارزن یا تخمینگر ساده شناخته می‌شود.

$$\varphi(u) = \begin{cases} 1 & |u_j| \leq \frac{1}{2}, \quad j = 1, \dots, d \\ 0 & \text{Otherwise} \end{cases}$$





پنجره پازن

تعداد کل نقاطی که در داخل ابرمکعب قرار می گیرند به صورت زیر است.

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

$$\varphi \left(\frac{x - x_i}{h_n} \right) = \begin{cases} 1 & |x - x_i| \leq \frac{h_n}{2}, \\ 0 & \text{Otherwise} \end{cases} \quad j = 1, \dots, d$$

تابع چگالی را می توان به صورت زیر در نظر گرفت

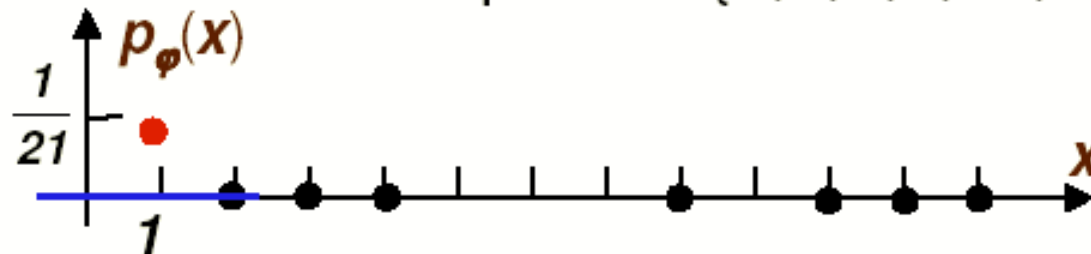
$$P_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{V_n} \varphi \left(\frac{x - x_i}{h_n} \right)$$

مقدار h_n عرض پنجره و $V_n = h_n^d$ است.

پنجره پارزن - مثال

$$p_{\varphi}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h^d} \varphi\left(\frac{x - x_i}{h}\right)$$

- Suppose we have 7 samples $D = \{2, 3, 4, 8, 10, 11, 12\}$



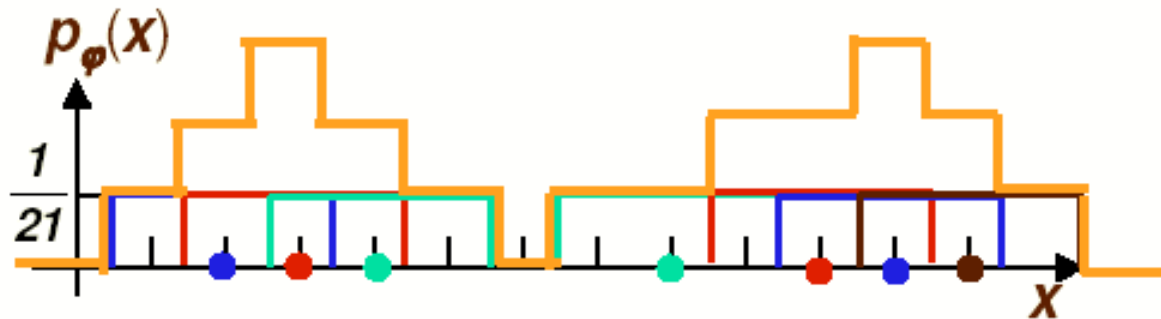
- Let window width $h=3$, estimate density at $x=1$

$$p_{\varphi}(1) = \frac{1}{7} \sum_{i=1}^7 \frac{1}{3} \varphi\left(\frac{1 - x_i}{3}\right) = \frac{1}{21} \left[\varphi\left(\frac{1-2}{3}\right) + \varphi\left(\frac{1-3}{3}\right) + \varphi\left(\frac{1-4}{3}\right) + \dots + \varphi\left(\frac{1-12}{3}\right) \right]$$

$$\left| -\frac{1}{3} \right| \leq 1/2 \quad \left| -\frac{2}{3} \right| > 1/2 \quad |-1| > 1/2 \quad \left| -\frac{11}{3} \right| > 1/2$$

$$p_{\varphi}(1) = \frac{1}{7} \sum_{i=1}^7 \frac{1}{3} \varphi\left(\frac{1 - x_i}{3}\right) = \frac{1}{21} [1 + 0 + 0 + \dots + 0] = \frac{1}{21}$$

- Let's come back to our example
 - 7 samples $\mathbf{D}=\{2,3,4,8,10,11,12\}$, $h=3$

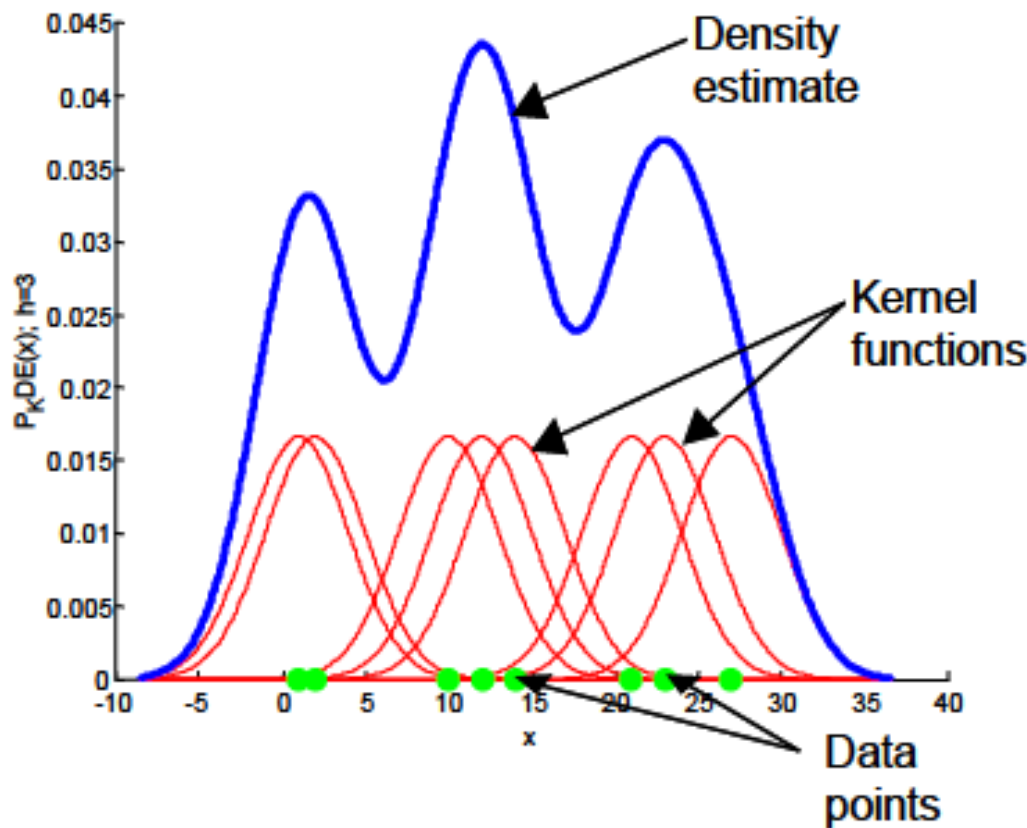


- To see what the function looks like, we need to generate 7 boxes and add them up
- The width is $h=3$ and the height, according to previous slide is

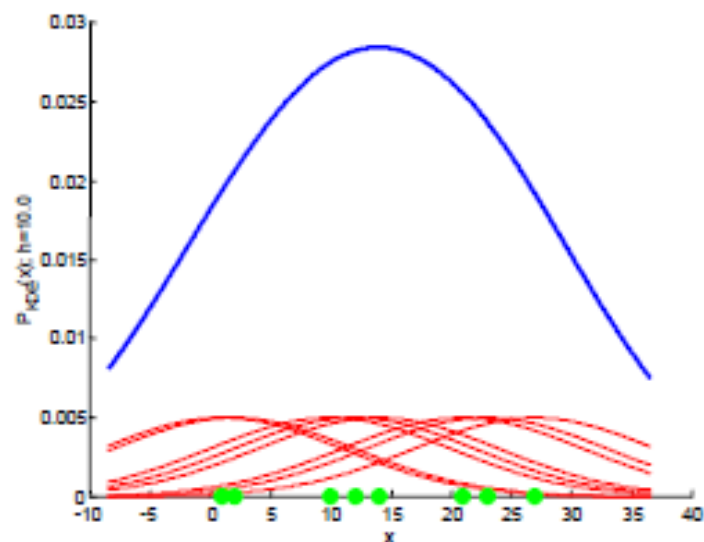
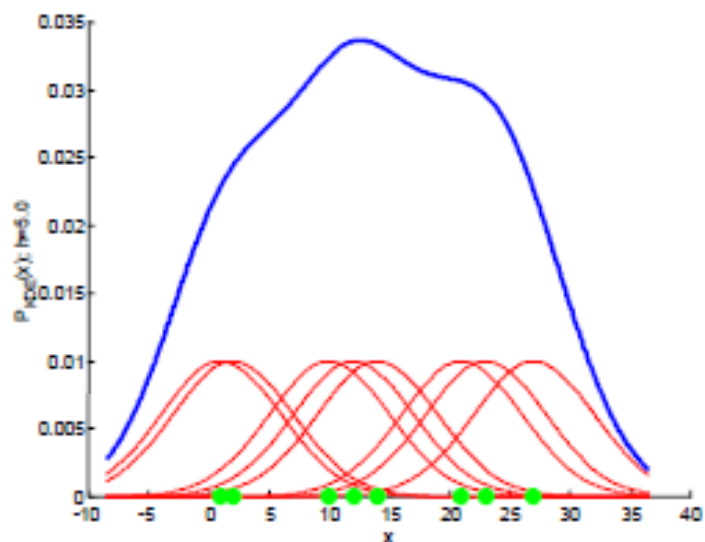
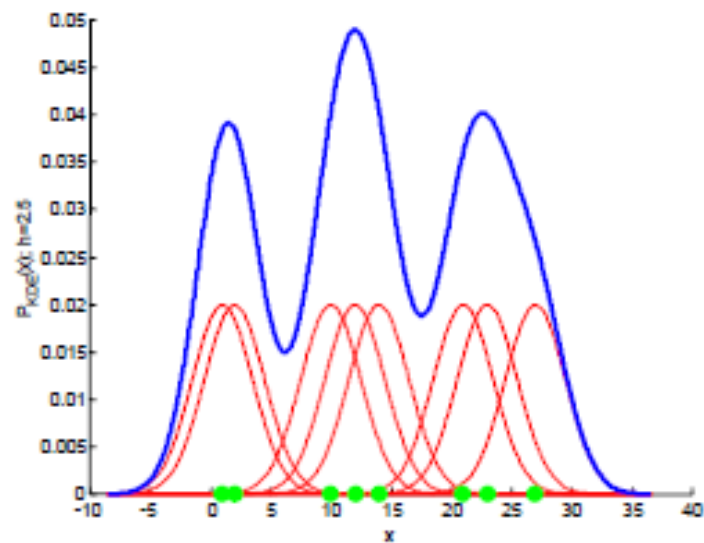
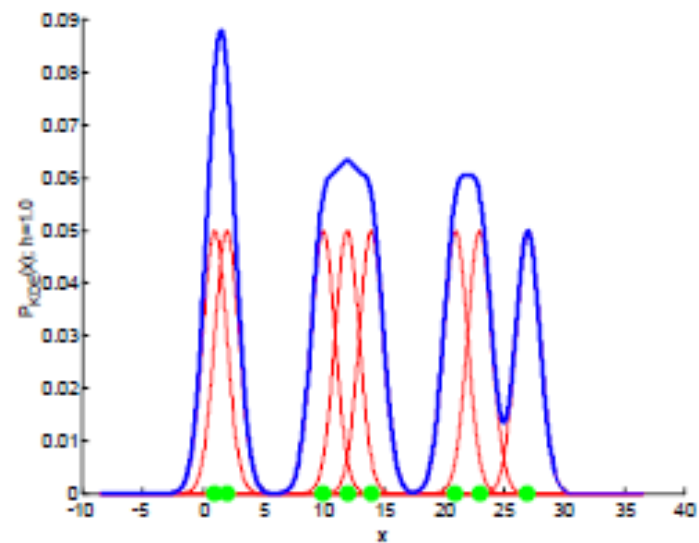
$$\frac{1}{nh^d} = \frac{1}{21}$$

پنجره نرم

○ به جای یک مکعب می‌توان یک برآمدگی نرم را در نظر گرفت که در این حالت تابع چگالی نرم‌تری حاصل می‌شود. در این حالت h پارامتر هموارسازی یا پهنای باند نیز نامیده می‌شود، که عرض برآمدگی را تعیین می‌کند با این قید که مساحت هر برآمدگی برابر یک است.



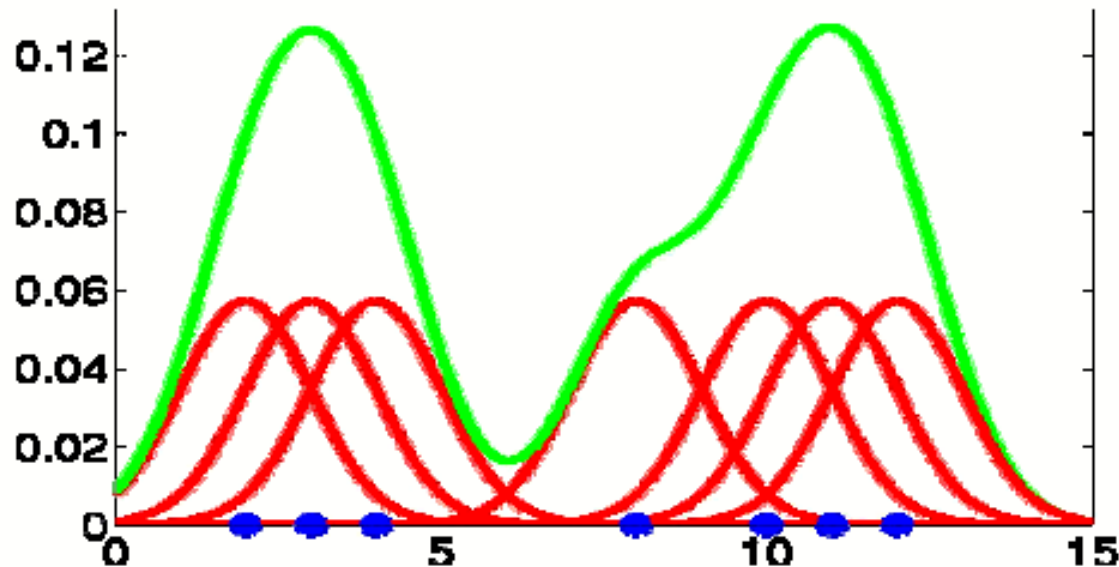
اثر پارامتر h



پنجره پارزن - مثال با پنجره نرم

- Let's come back to our example
 - 7 samples $D=\{2,3,4,8,10,11,12\}$, $h=1$

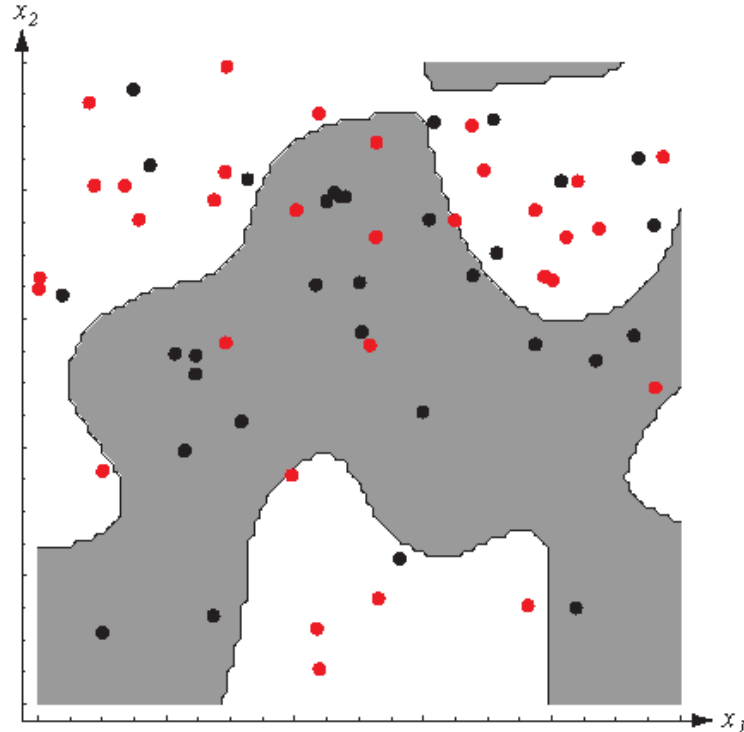
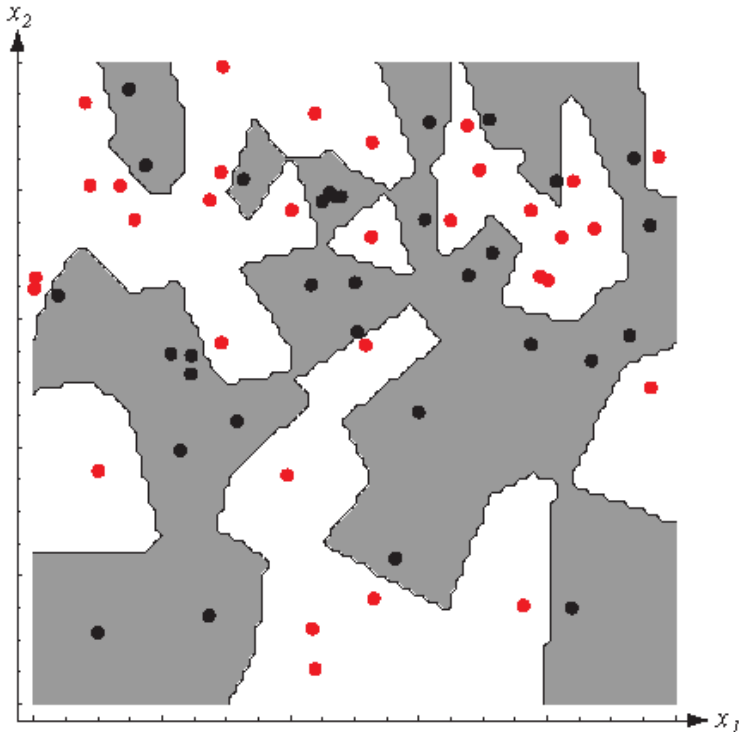
$$p_{\varphi}(x) = \frac{1}{7} \sum_{i=1}^7 \varphi(x - x_i)$$



- $p_{\varphi}(x)$ is the sum of 7 Gaussians, each centered at one of the sample points, and each scaled by $1/7$

پنجره پارزن

- تابع چگالی تخمین زده شده با استفاده از پنجره‌های پارزن را می‌توان به همراه قانون تصمیم بیزی برای طبقه بندی استفاده کرد.
- خطای آموزشی را می‌توان با کوچک کردن عرض پنجره کم کرد. اما باید مراقب تعمیم بود.
- شکل زیر مرزهای تصمیم در یک تصویر فضای دو بعدی را نشان می‌دهد. سمت چپ: از یک پنجره با عرض کوچک استفاده می‌کند، سمت راست: از عرض پنجره بزرگتری استفاده می‌کند.



معایب روش‌های مبتنی بر کرنل

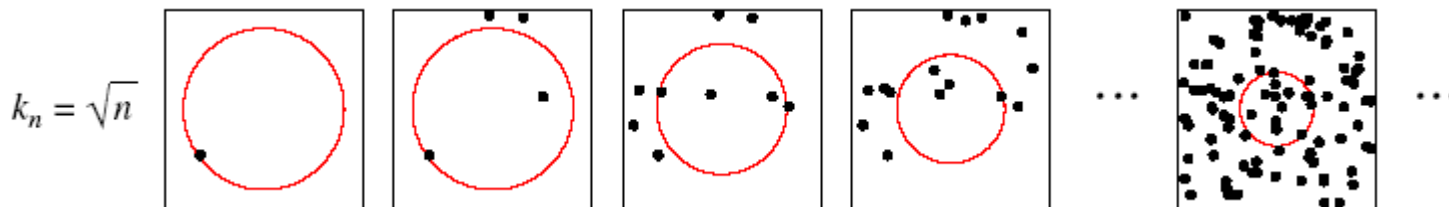
- نیاز به تعداد زیادی نمونه داریم.
- تمام نمونه‌ها باید ذخیره شوند.
- اگر تعداد نقاط داده زیاد باشد، ارزیابی چگالی می‌تواند بسیار کند باشد.

راه حل پیشنهادی: پس از محاسبه چگالی، این فضا با هسته‌های کمتری تقریب زده شود. (مخلوطی از گاوس‌ها)

K نزدیک‌ترین همسایگی

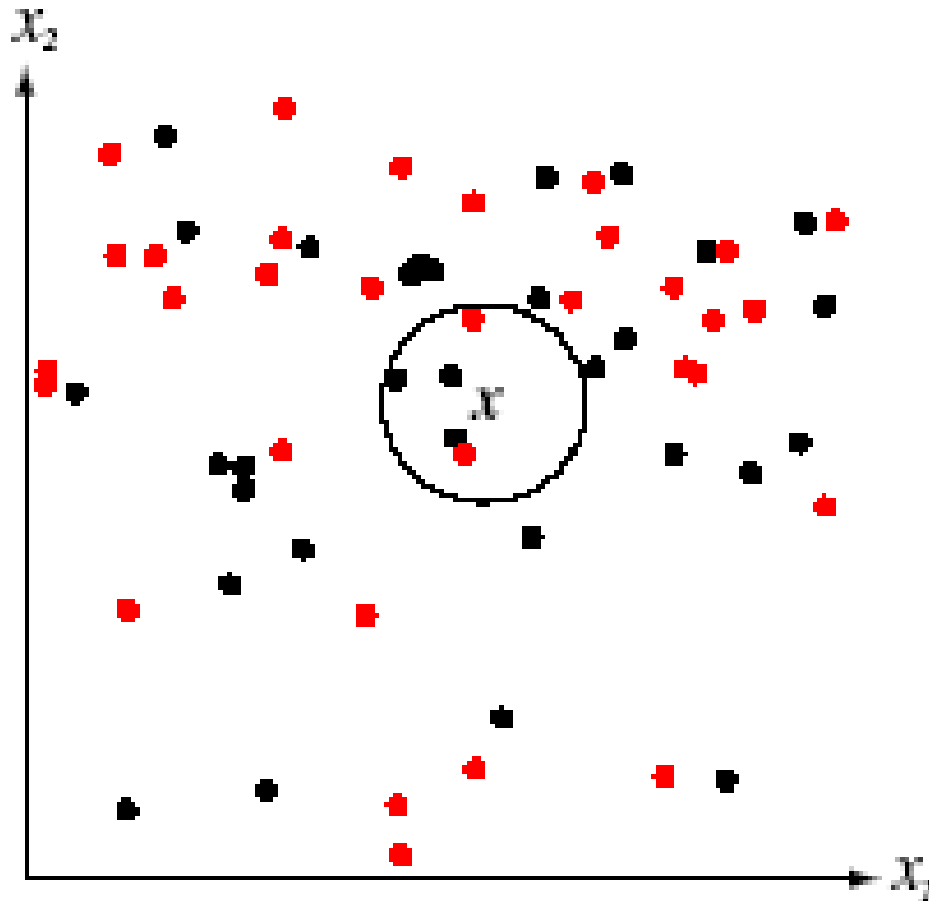
K-NEAREST NEIGHBORS

- اندازه مناسب پنجره یک چالش در پنجره پارزن است.
- یک راه حل این است که حجم تخمین، تابعی از داده‌های آموزشی باشد، نه یک تابع دلخواه.
- برای تخمین کلاس در مکان x ، فضای بررسی را آنقدر بزرگ کنیم تا k نمونه در آن قرار گیرد و با این تعداد همسایگی کلاس‌بندی صورت گیرد. مقدار k تابعی از کل نمونه‌ها خواهد بود.
- اگر چگالی داده‌ها نزدیک به x زیاد باشد، فضای بررسی کوچک خواهد بود، و اگر چگالی کم باشد، فضا بزرگ می‌شود.

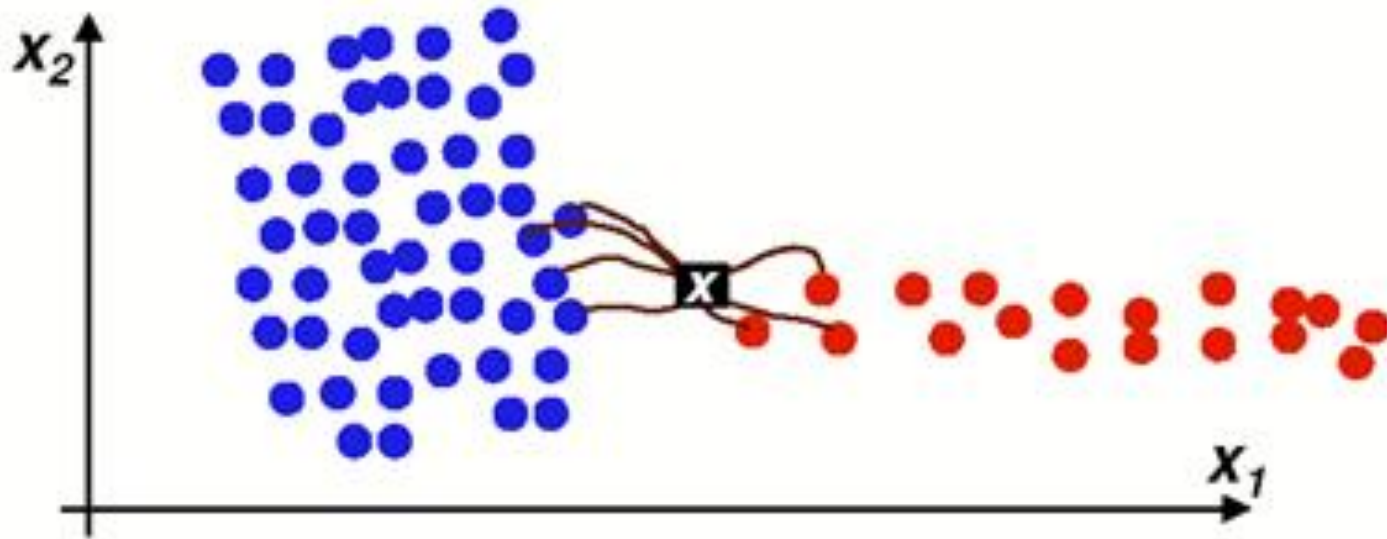


K نزدیک‌ترین همسایگی - مثال

در شکل روبرو x متعلق به کلاس سیاه است. با فرض $k = 5$

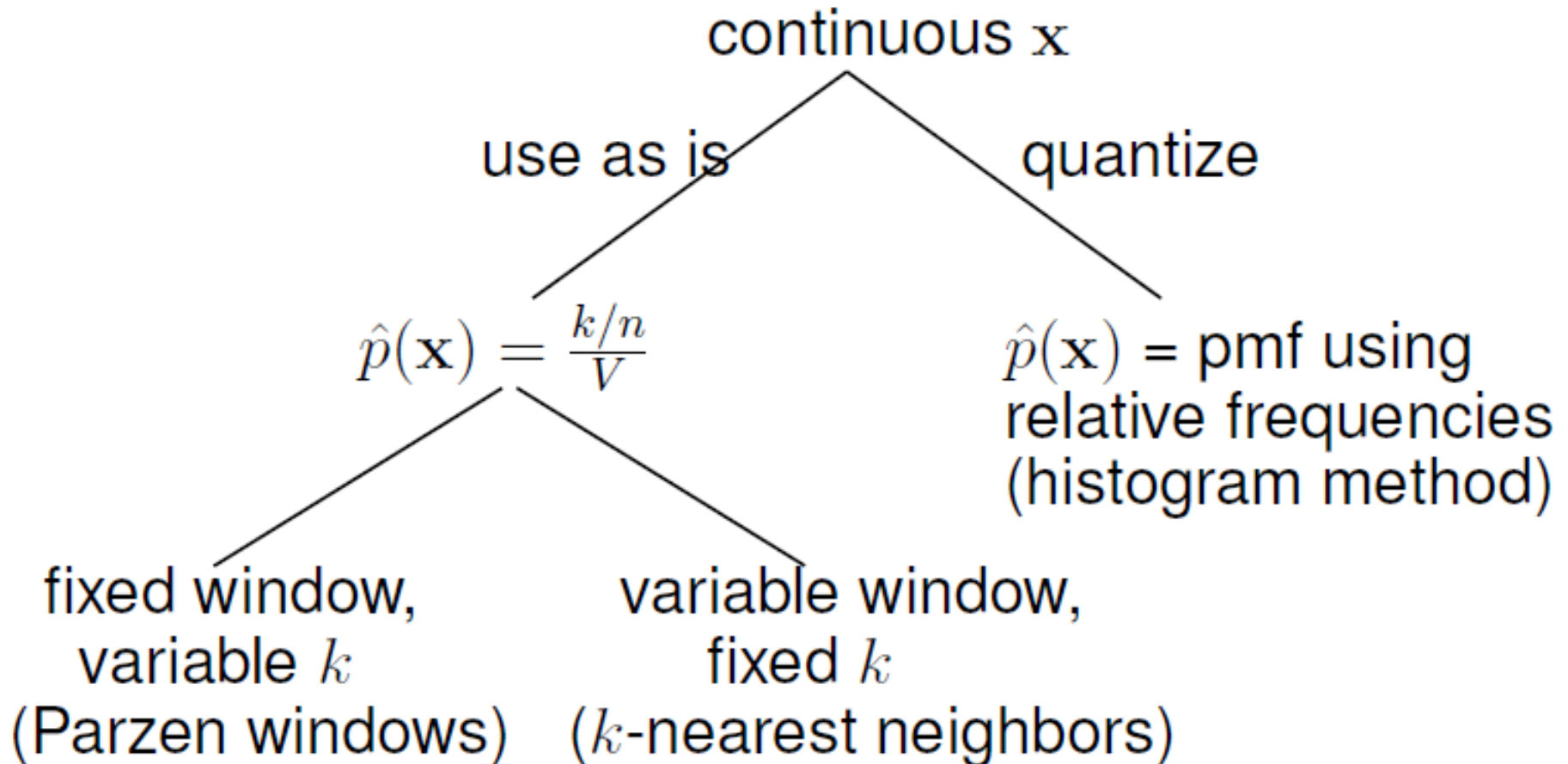


چالش در نزدیک‌ترین همسایگی: مقدار k



- For $k = 1, \dots, 5$ point x gets classified correctly
 - red class
- For larger k classification of x is wrong
 - blue class

روش‌های بدون پارامتر



روش‌های بدون پارامتر

○ مزایا

- هیچ فرضی در مورد توزیع‌های پیش‌رو لازم نیست (عمومی).
- با تعداد نمونه کافی، به هر تابع چگالی دلخواهی (پیچیده‌ای) همگرا می‌شود.

○ معایب

- تعداد نمونه‌های مورد نیاز ممکن است بسیار زیاد باشد (تعداد با ابعاد این ویژگی به طور تصاعدی رشد می‌کند).
- ممکن است حافظه و زمان محاسبات زیادی لازم داشته باشد.

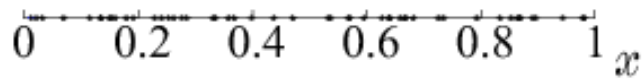
مساله بعد بالا

CURSE OF DIMENSIONALITY

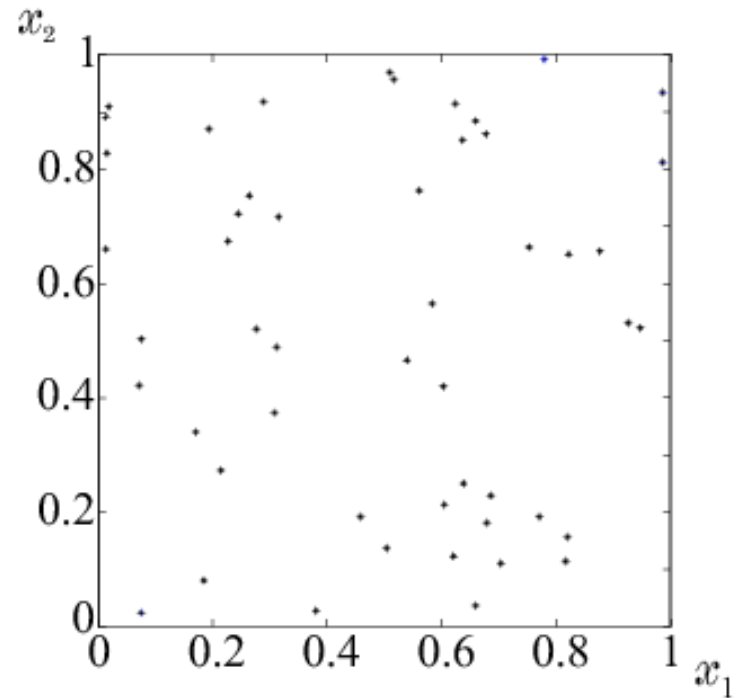
- در تمام روش‌ها که تا کنون شاهد بودیم؛ هر چه تعداد نقاط، N بیشتر باشد، تخمین بهتری حاصل می‌شود.
- اگر در فضای یک بعدی N نقطه کافی باشد (برای تخمین خوب)، در فضای دو بعدی به N^2 و در فضای l بعدی، به N^l نقطه نیاز دارد.
- افزایش تصاعدی در تعداد نقاط ضروری به مساله بعد بالا (نفرین ابعاد) شناخته می‌شود. این مشکل عمده‌ای است که در فضاهای با ابعاد بالا با آن مواجه هستیم.

مساله بعد بالا - مثال

CURSE OF DIMENSIONALITY



(a)



(b)

NAIVE – BAYES CLASSIFIER

- در یک مساله l بعدی، تعداد نقطه لازم N^l است. با فرض استقلال ویژگی ها داریم:

$$p(\underline{x} | \omega_i) = \prod_{j=1}^{\ell} p(x_j | \omega_i)$$

- در این مورد، تقریباً به N نقطه برای هر تابع چگالی نیاز داریم. بنابراین، تعدادی نقطه از مرتبه N^l کافی است.
- به نظر می رسد که طبقه بندی کننده نایوبیز حتی در مواردی که فرض استقلال نقض می شود، به خوبی کار می کند.

تمرین

۱- با درد نظر گرفتن پنجره پارزن، تابع چگالی احتمال داده‌های زیر را رسم نمایید. کرنل را یک مثلث با قاعده دو و ارتفاع یک در نظر بگیرید.

1, 4, 3, 4, 0, 2, 4, 2

۲- فاصله اقلیدسی و ماهالانوبیس دو نقطه (۱ و ۲) و (۳ و ۶) را بیابید. اگر این دو نقطه را مرکز دو کلاس با کواریانس یکسان داده شده در نظر بگیریم. در مورد مرز جدا کننده توضیح دهید.

$$\Sigma = \begin{bmatrix} 4 & 1 \\ 1 & 2 \end{bmatrix}$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

۳- اگر ریسک کلاس‌بندی نادرست داده‌های با تابع چگالی مثلث دو برابر داده‌ها با تابع چگالی یکنواخت باشد. مرز بهینه را بیابید. احتمال پیشین دو کلاس یکسان است.

