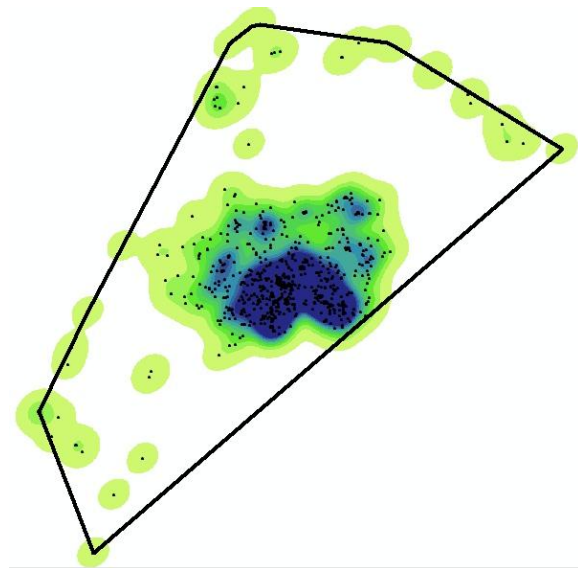


Estimacija GRV pomoću kernela (eng. *kernel density estimation – KDE*)

- Neparametarska estimacija gustine verovatnoće
- Histogrami
- KDE pomoću pravougaonog kernela
- Glatki kerneli
- Separabilni kerneli
- Naivni Bayesov klasifikator

Neparametarska estimacija gustine verovatnoće

- Bayesova teorija odlučivanja je formalno definisala problem određivanja regiona odlučivanja i klasifikatora *uz pretpostavku da je raspodela poznata*
- Najčešće ne poznajemo pravu gustinu raspodele verovatnoće već se ona mora proceniti na osnovu eksperimentalnih podataka
 - Estimacija parametara
 - Neparametarska estimacija
- Neparametarska estimacija gustine raspodele verovatnoće
 - Nema pretpostavke o parametarskom obliku gustine raspodele verovatnoće
 - Postoje i metode čiji je cilj direktna klasifikacija bez eksplicitne estimacije gustine raspodele verovatnoće



Histogram

■ Najjednostavniji metod neparametarske estimacije

- Prostor uzoraka deli se na ćelije
- Verovatnoća da će uzorak upasti u neku ćeliju aproksimira se relativnim brojem uzoraka iz skupa za obuku koji pripadaju posmatranoj ćeliji:

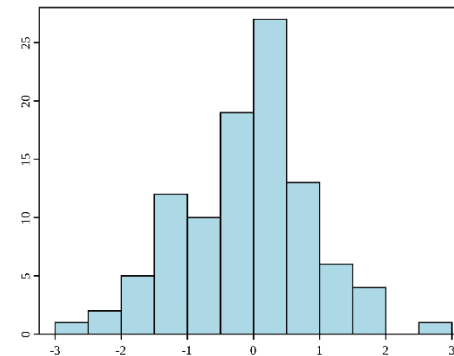
$$p \approx \frac{k_N}{N}$$

gde je k_N broj uzoraka u određenoj ćeliji a N ukupan broj uzoraka

- Procenjena gustina raspodele verovatnoće je konstantna unutar ćelije širine h :

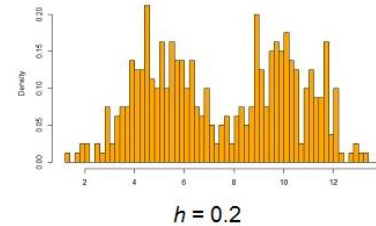
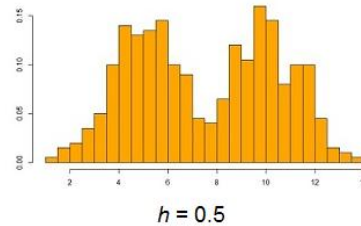
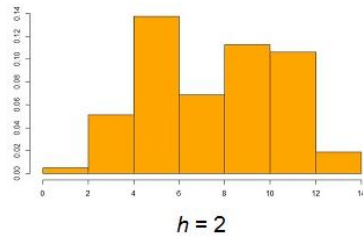
$$\hat{p}_H(x) \approx \frac{1}{h} \cdot \frac{k_N}{N}, \quad |x - \hat{x}| \leq \frac{h}{2}$$

i konvergira ka pravoj gustini raspodele ako $h \rightarrow 0$, $k_N \rightarrow \infty$ i $k_N/N \rightarrow 0$ (što su logični uslovi)

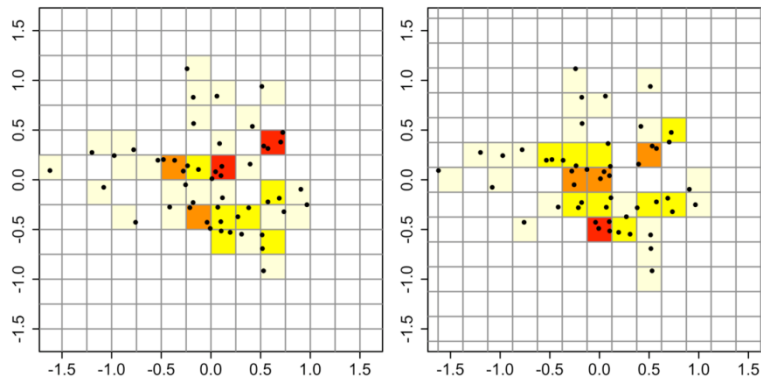


Nedostaci histograma

- Izrazita zavisnost od veličine ćelije
 - Problem određivanja optimalne širine ćelije



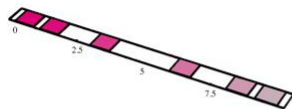
- Zavisnost od položaja prve ćelije



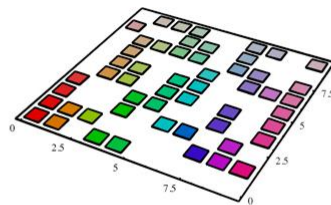
- Kod višedimenzijskih obeležja javlja se i zavisnost od orijentacije ćelije

Nedostaci histograma

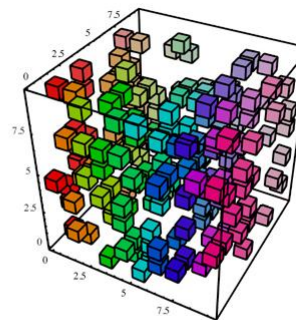
- Diskontinuiteti u estimiranoj gustini raspodele verovatnoće
 - Artefakt histograma kao metode
- Problem kod prostora visoke dimenzionalnosti (eng. *the curse of dimensionality*)
 - Potreban je ogroman broj uzoraka da većina ćelija ne bi ostala prazna
 - Da bi se postigla podjednaka popunjenost prostora u d dimenzija potrebno je N^d uzoraka



1-D



2-D



3-D

- Zbog svih navedenih nedostataka primena histograma ograničena je uglavnom na brzu vizuelizaciju podataka (u jednoj ili dve dimenzije)

Opšta formulacija neparametarske estimacije GRV

- Verovatnoća da će vektor \mathbf{x} , izvučen iz raspodele $p(\mathbf{x})$, upasti u određeni region R uzoračkog prostora u opštem slučaju iznosi:

$$P = \int_R p(\mathbf{x}) d\mathbf{x}$$

- Neka je N vektora $\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}\}$ izvučeno iz raspodele. Verovatnoća da se k od ovih N vektora nađe u R data je *binomnom raspodelom*:

$$P[k] = \binom{N}{k} P^k (1-P)^{N-k}$$

- Ako se količnik k/N posmatra kao slučajna promenljiva, može se pokazati da su njegova srednja vrednost i varijansa:

$$E\left(\frac{k}{N}\right) = P, \quad \text{Var}\left(\frac{k}{N}\right) = E\left(\left(\frac{k}{N} - P\right)^2\right) = \frac{P(1-P)}{N}$$

- Dakle, kada $N \rightarrow \infty$, raspodela postaje oštija (varijansa teži nuli), pa se može očekivati dobra estimacija verovatnoće P kao količnika broja uzoraka koji upadaju u region R i ukupnog broja uzoraka

Opšta formulacija neparametarske estimacije GRV

- Uz pretpostavku da je R dovoljno malo tako da se $p(\mathbf{x})$ ne menja značajno unutar njega, važi:

$$P = \int_R p(\mathbf{x}) d\mathbf{x} \approx p(\mathbf{x}) \cdot V$$

gde je V zapremina regiona R , odakle se dobija:

$$\hat{p}(\mathbf{x}) = \frac{P}{V} = \frac{k_N}{NV}$$

- Ova estimacija postaje sve tačnija sa porastom N i smanjenjem V
- U praksi je veličina skupa uzoraka N fiksna
 - U cilju poboljšanja tačnosti estimacije $p(\mathbf{x})$, može se pustiti da se V približava nuli, ali bi onda region R postao tako mali da ne bi obuhvatao nijedan uzorak
 - U praksi se mora naći kompromisna vrednost za V , koja mora biti:
 - dovoljno velika da R obuhvati dovoljan broj uzoraka raspodele
 - dovoljno mala da bi važila pretpostavka da je $p(\mathbf{x})$ konstantno unutar R

Opšta formulacija neparametarske estimacije GRV

- Opšti izraz za neparametarsku estimaciju gustine verovatnoće je:

$$\hat{p}(\mathbf{x}) = \frac{k_N}{NV}$$

V – zapremina koja obuhvata \mathbf{x}

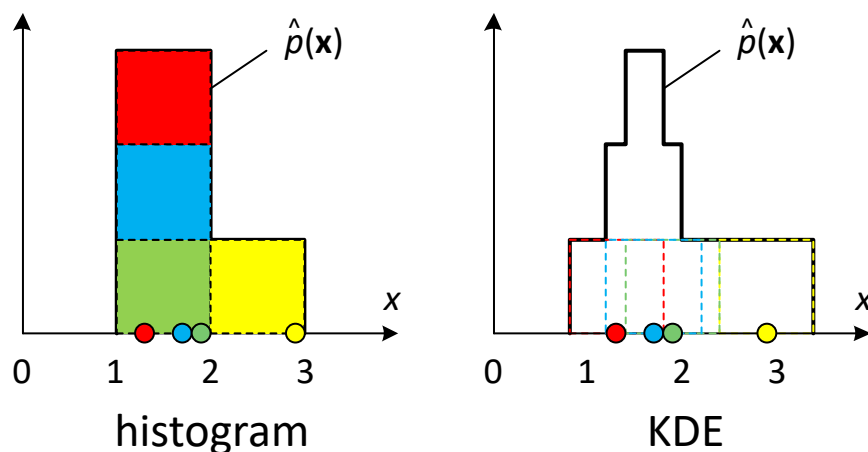
N – ukupan broj uzoraka

k_N – broj uzoraka unutar V

- U praktičnoj primeni ovog izraza postoje dva osnovna pristupa
 - Fiksirati zapreminu V i odrediti k_N na osnovu skupa uzoraka – što se naziva metoda **estimacije gustine raspodele verovatnoće pomoću kernela** (eng. *kernel density estimation* – KDE)
 - Fiksirati vrednost k_N i odrediti odgovarajuću zapreminu oko tačke estimacije, čime se bavi metoda **k najbližih suseda** (eng. *k nearest neighbors* – kNN)
- Estimacije $p(\mathbf{x})$ dobijene bilo kojom od ove dve metode (KDE i kNN) konvergiraju ka stvarnoj gustini raspodele verovatnoće kada $N \rightarrow \infty$, $V \rightarrow 0$, $k_N \rightarrow \infty$ i $k_N/N \rightarrow 0$

Estimacija GRV pomoću kernela (KDE)

- Svaki pojedinačni uzorak daje određeni doprinos ukupnoj estimiranoj GRV
- Procena GRV predstavljena je u vidu sume pojedinačnih tzv. *kernel* funkcija lociranih na mestima gde se nalaze pojedinačni uzorci
- Histogram se može zamisliti kao KDE kod koje je svaki uzorak implicitno premešten u centar ćelije u kojoj se nalazi (što ne bi bilo opravdano)



- Na ovaj način prevazilazi se zavisnost histograma od položaja početne ćelije
 - Prikazani primer odnosi se na KDE korišćenjem tzv. Parzenovog (pravougaonog) prozora, gde je doprinos uzorka ukupnoj GRV isti na čitavom intervalu širine h

Parzenov prozor (kernel)

- Neka region R predstavlja D -dimenzionalnu hiperkocku stranice dužine h , sa centrom u tački estimacije \mathbf{x}
 - Zapremina takve kocke je $V = h^D$
- Ako se sa $K(\mathbf{u})$ označi jedinična hiperkocka centrirana oko koordinatnog početka (Parzenov prozor):

$$K(\mathbf{u}) = \begin{cases} 1, & |u_j| < 1/2, \forall j \in \{1, \dots, D\} \\ 0, & \text{drugde} \end{cases}$$

uzorak $\mathbf{x}^{(n)}$ pripada hiperkocki stranice h centriranoj oko \mathbf{x} ako i samo ako važi:

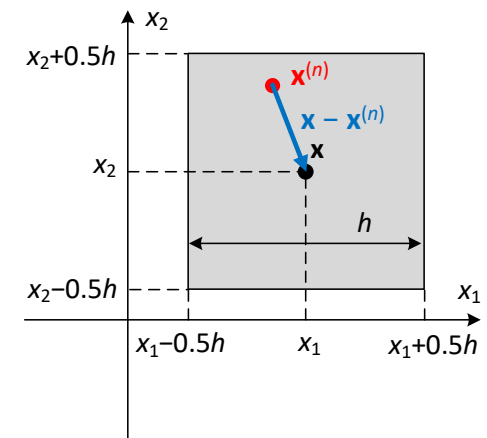
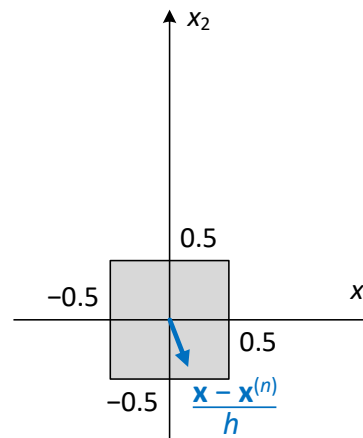
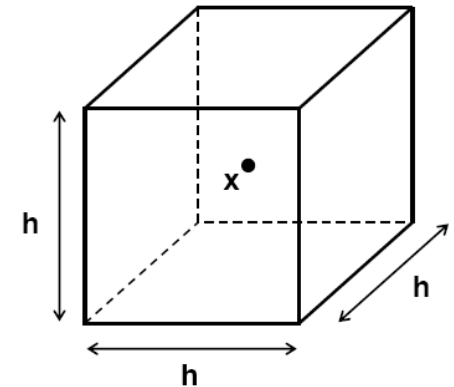
$$K\left(\frac{\mathbf{x} - \mathbf{x}^{(n)}}{h}\right) = 1$$

- Broj uzoraka unutar te hiperkocke je:

$$k_N = \sum_{n=1}^N K\left(\frac{\mathbf{x} - \mathbf{x}^{(n)}}{h}\right)$$

a procena GRV jednaka je:

$$\hat{p}_{KDE}(\mathbf{x}) = \frac{1}{N \cdot h^D} \sum_{n=1}^N K\left(\frac{\mathbf{x} - \mathbf{x}^{(n)}}{h}\right)$$



Parzenov prozor (kernel)

- Neka region R predstavlja D -dimenzionalnu hiperkocku stranice dužine h , sa centrom u tački estimacije \mathbf{x}
 - Zapremina takve kocke je $V = h^D$
- Ako se sa $K(\mathbf{u})$ označi jedinična hiperkocka centrirana oko koordinatnog početka (Parzenov prozor):

$$K(\mathbf{u}) = \begin{cases} 1, & |u_j| < 1/2, \forall j \in \{1, \dots, D\} \\ 0, & \text{drugde} \end{cases}$$

uzorak $\mathbf{x}^{(n)}$ pripada hiperkocki stranice h centriranoj oko \mathbf{x} ako i samo ako važi:

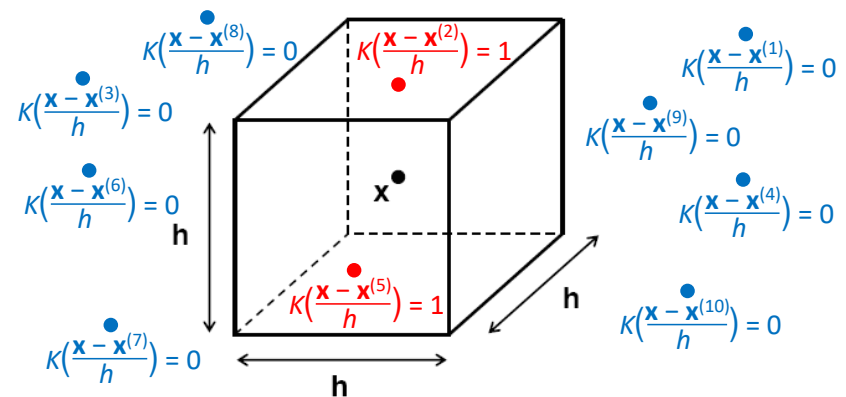
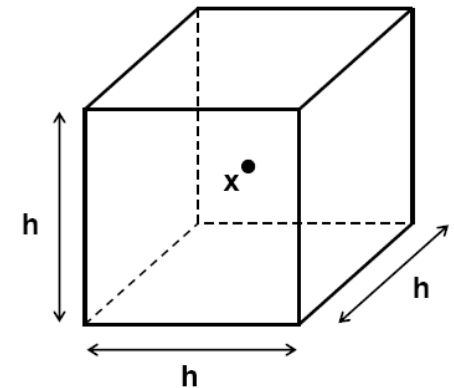
$$K\left(\frac{\mathbf{x} - \mathbf{x}^{(n)}}{h}\right) = 1$$

- Broj uzoraka unutar te hiperkocke je:

$$k_N = \sum_{n=1}^N K\left(\frac{\mathbf{x} - \mathbf{x}^{(n)}}{h}\right)$$

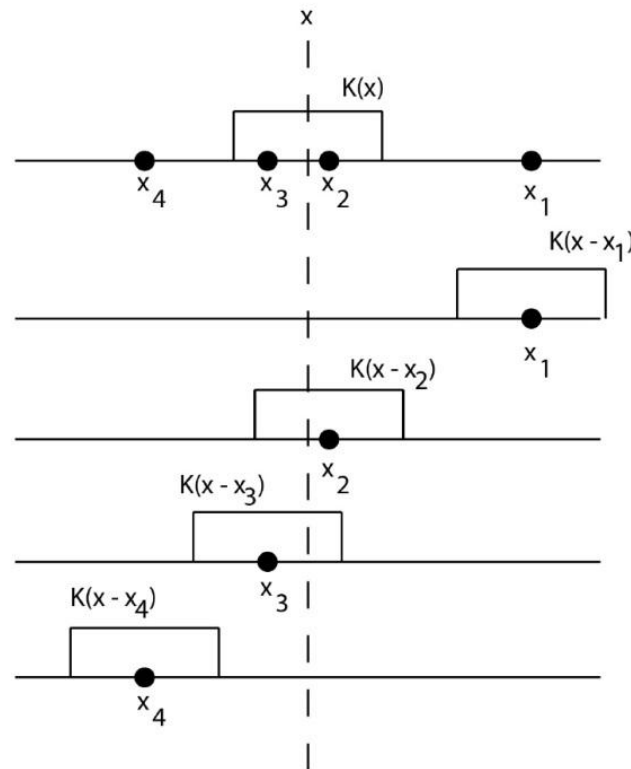
a procena GRV jednaka je:

$$\hat{p}_{KDE}(\mathbf{x}) = \frac{1}{N \cdot h^D} \sum_{n=1}^N K\left(\frac{\mathbf{x} - \mathbf{x}^{(n)}}{h}\right)$$



Parzenov prozor (kernel)

- Moguća je i alternativna interpretacija
 - Broj uzoraka koji upadaju u hiperkocku centriranu oko proizvoljne tačke \mathbf{x} ujedno predstavlja broj hiperkocka centriranih oko pojedinačnih uzoraka koje se preklapaju u tački \mathbf{x}



Očekivanje KDE estimacije

- Očekivana vrednost KDE estimacije je:

$$\begin{aligned} E(\hat{p}_{KDE}(\mathbf{x})) &= \frac{E(k_N)}{N \cdot h^D} = \frac{1}{N \cdot h^D} \sum_{n=1}^N E\left(K\left(\frac{\mathbf{x} - \mathbf{x}^{(n)}}{h}\right)\right) \\ &= \frac{1}{h^D} E\left(K\left(\frac{\mathbf{x} - \mathbf{x}^{(n)}}{h}\right)\right) \\ &= \frac{1}{h^D} \int K\left(\frac{\mathbf{x} - \mathbf{x}'}{h}\right) p(\mathbf{x}') d\mathbf{x}' \\ &= \frac{1}{h^D} K\left(\frac{\mathbf{x}}{h}\right) * p(\mathbf{x}) \end{aligned}$$

Konvolucija (podsećanje)

$$f(x) * g(x) = \int_x f(\xi) * g(x - \xi) d\xi$$

$$f(x) * g(x) = g(x) * f(x)$$

$$f(x) * A\delta(x - x_0) = Af(x - x_0)$$

- KDE estimacija je pristrasna, jer $E(\hat{p}_{KDE}(\mathbf{x})) \neq p(\mathbf{x})$

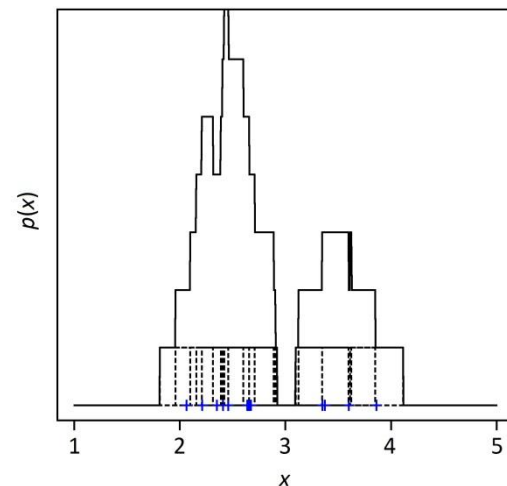
Veza stvarne GRV i njene procene

- Očekivanje KDE estimacije je konvolucija stvarne GRV $p(\mathbf{x})$ i kernela
 - Konvolucija izaziva ublaženje naglih skokova u stvarnoj gustini raspodele, u meri određenoj vrstom i širinom kernela
 - Za $h \rightarrow 0$ (u 1-D slučaju), kernel se približava Diracovom δ -impulsu:

$$\delta(x) = \begin{cases} \infty, & x = 0 \\ 0, & x \neq 0 \end{cases} \quad \int_{-\infty}^{\infty} \delta(x) dx = 1$$

i estimacija postaje nepristrasna, ali se u praksi, za konačno N , estimacija pretvara u skup impulsa lociranih u tačkama skupa za obuku (što je loše)

- U opštem slučaju, za konačno N , $p(\mathbf{x})$ se estimira kao suma konačnog broja prekidnih funkcija, pa je i sama estimacija prekidna funkcija
 - Ideja: korišćenje neprekidnih (glatkih) funkcija kao kernela

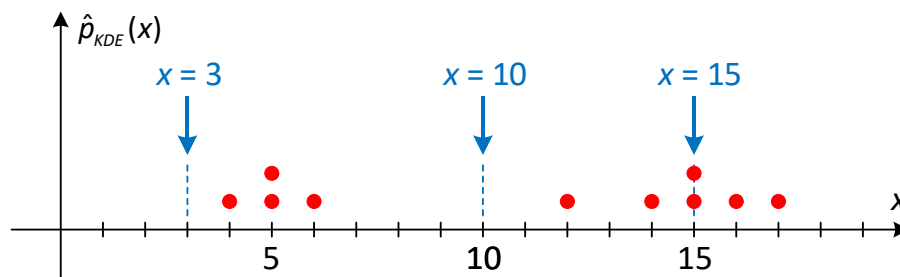


Primer

- Na osnovu datog skupa uzoraka $X = \{x^{(1)}, x^{(2)}, \dots, x^{(N)}\} = \{4, 5, 5, 6, 12, 14, 15, 15, 16, 17\}$ estimirati gustinu raspodele verovatnoće $p(x)$ u tačkama $x = 3, 10, 15$ pomoću Parzenovih prozora širine $h = 4$.

- Rešenje:

$$\hat{p}_{KDE}(x) = \frac{1}{N \cdot h^D} \sum_{n=1}^N K\left(\frac{x - x^{(n)}}{h}\right)$$



$$\hat{p}_{KDE}(x)|_{x=3} = \frac{1}{10 \cdot 4^1} \left[\underbrace{K\left(\frac{3-4}{4}\right)}_{-1/4} + \underbrace{K\left(\frac{3-5}{4}\right)}_{-1/2} + \underbrace{K\left(\frac{3-5}{4}\right)}_{-1/2} + \underbrace{K\left(\frac{3-6}{4}\right)}_{-3/4} + \dots + \underbrace{K\left(\frac{3-17}{4}\right)}_{-14/4} \right] =$$

$$= \frac{1}{40} [1 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0] = \frac{1}{40} = 0.025$$

$$\hat{p}_{KDE}(x)|_{x=10} = \frac{1}{40} [0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0] = 0$$

$$\hat{p}_{KDE}(x)|_{x=15} = \frac{1}{40} [0 + 0 + 0 + 0 + 0 + 0 + 1 + 1 + 1 + 1 + 0] = \frac{4}{40} = 0.1$$

Glatki kerneli

- Izbor glatkog kernela umesto Parzenovog prozora:
 - Otklanja problem diskontinuiteta u proceni GRV
 - Omogućuje da se doprinosi različitih tačaka u okolini uzorka različito ponderišu
- Kernel $K(\mathbf{x})$ mora zadovoljiti uslov:

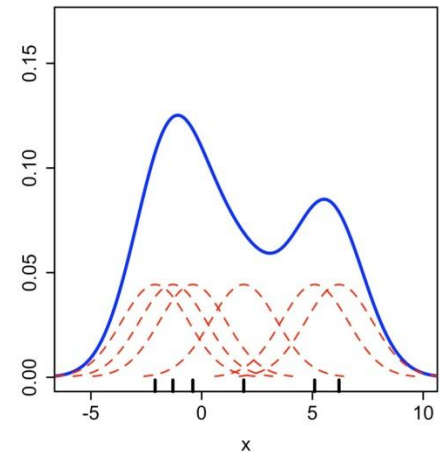
$$\int_{R^D} K(\mathbf{x}) d\mathbf{x} = 1,$$

i obično se koriste glatke, radijalno simetrične i unimodalne funkcije kao što je Gaussova:

$$K(\mathbf{x}) = \frac{1}{(2\pi)^{D/2}} e^{-\frac{1}{2}\mathbf{x}^T\mathbf{x}}$$

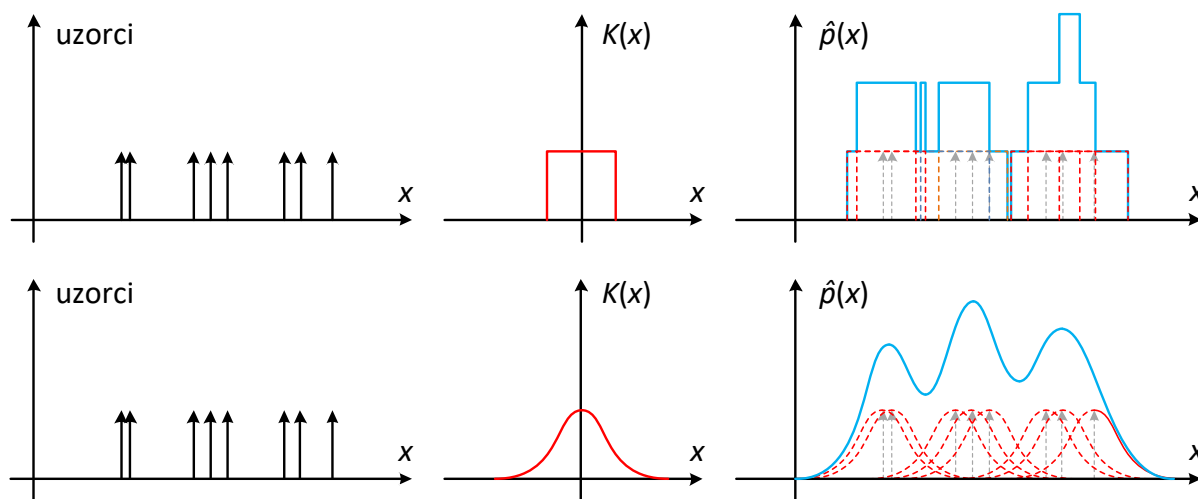
- Izraz za procenu gustine verovatnoće ostaje nepromenjen:

$$\hat{p}_{KDE}(\mathbf{x}) = \frac{1}{N \cdot h^D} \sum_{n=1}^N K\left(\frac{\mathbf{x} - \mathbf{x}^{(n)}}{h}\right)$$



Predstavljanje KDE estimacije preko konvolucije

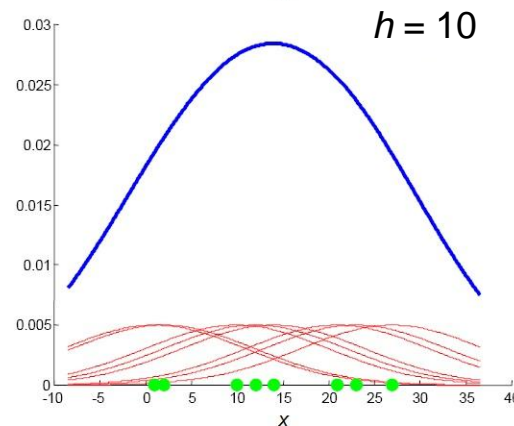
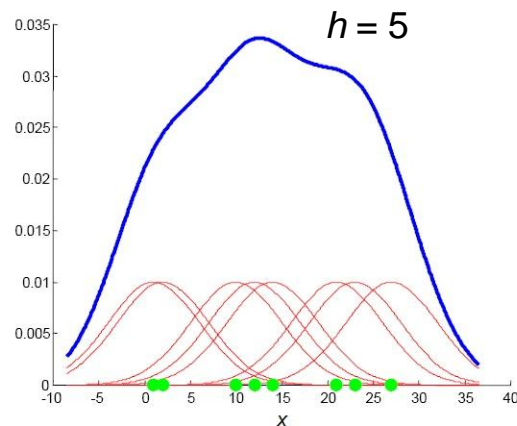
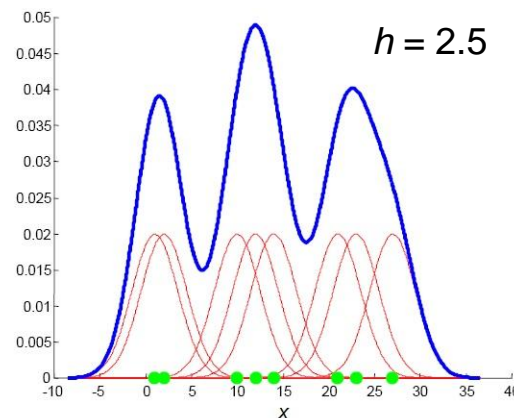
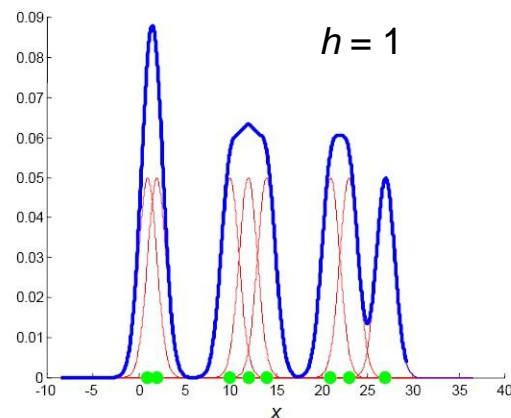
- Konvolucija u teoriji signala i sistema predstavlja vezu između pobude i odziva linearnog vremenski nepromenljivog sistema (filtra)
 - U slučaju Parzenovog prozora, $\hat{p}_{KDE}(\mathbf{x})$ se dobija kao odziv filtra sa pravougaonim impulsnim odzivom na pobudu u vidu niza δ -impulsa
 - U slučaju glatkog kernela, $\hat{p}_{KDE}(\mathbf{x})$ se dobija kao odziv filtra sa glatkim impulsnim odzivom na istu pobudu



- Izbor kernela određuje oblik modova (izbočina) u estimiranoj raspodeli
- Širina modova određena je širinom kernela (eng. *kernel bandwidth*)

Izbor širine kernela

- Izbor širine kernela je od suštinskog značaja za primenu KDE metode
 - Preuzak kernel daje estimaciju sa dosta (lažnih) modova, koja je teška za interpretaciju
 - Preširok kernel suviše izravnavava estimaciju i skriva strukturu stvarne raspodele



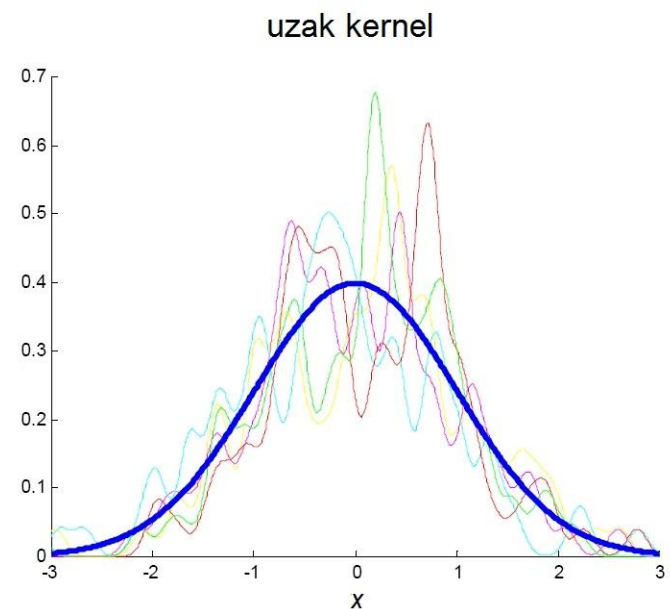
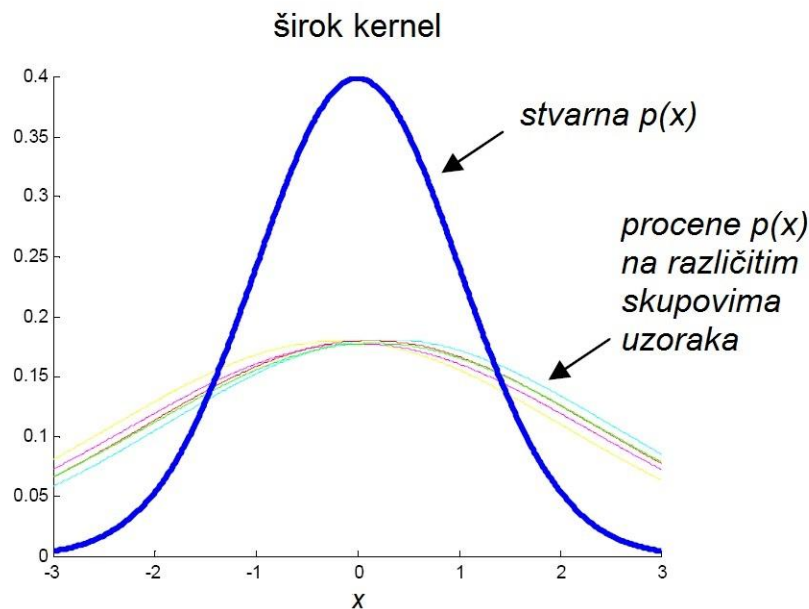
Uticaj izbora širine kernela na procenu GRV

■ Izbor širokog kernela

- smanjuje varijansu jer su procene na različitim skupovima uzoraka međusobno slične
- povećava pristrasnost jer prosek svih procena $\hat{p}_{KDE}(\mathbf{x})$ manje liči na stvarnu GRV

■ Izbor uskog kernela

- smanjuje pristrasnost jer prosek svih procena $\hat{p}_{KDE}(\mathbf{x})$ više liči na stvarnu GRV
- povećava varijansu jer su procene na različitim skupovima uzoraka značajno različite



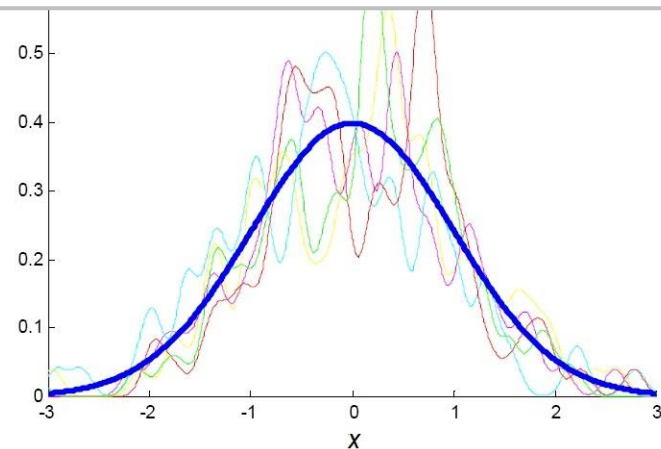
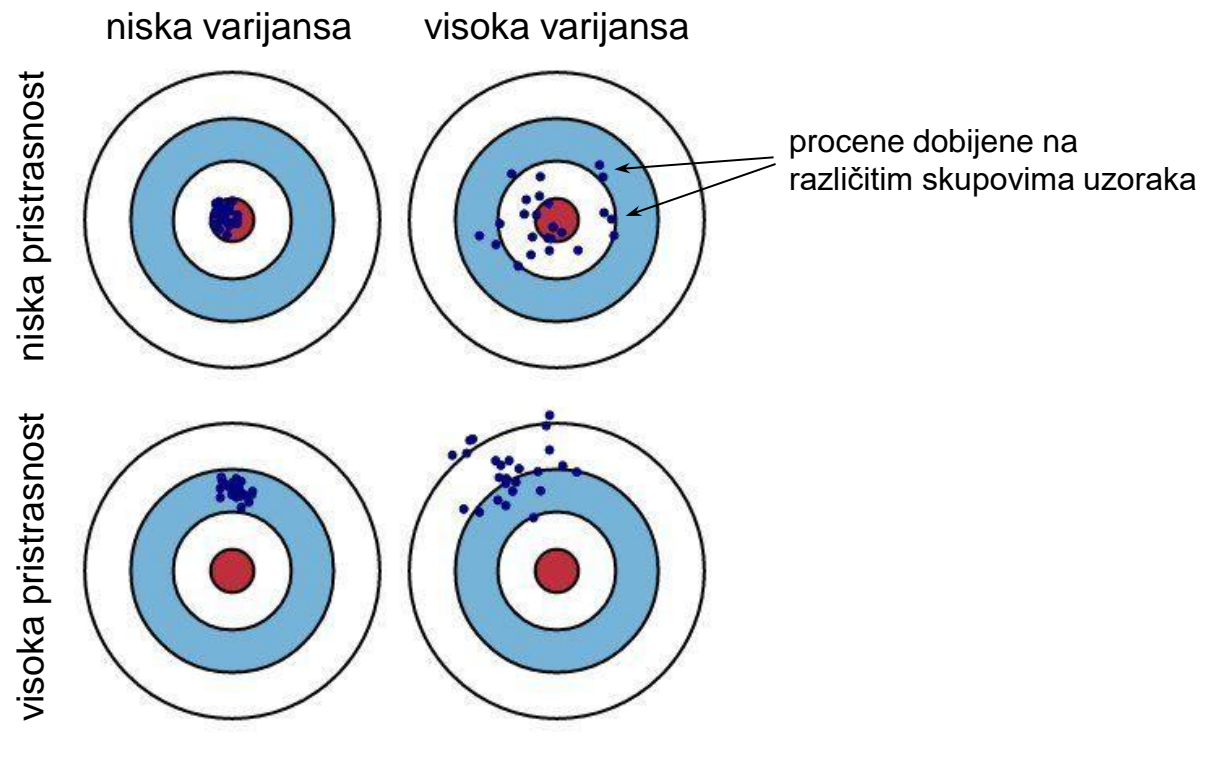
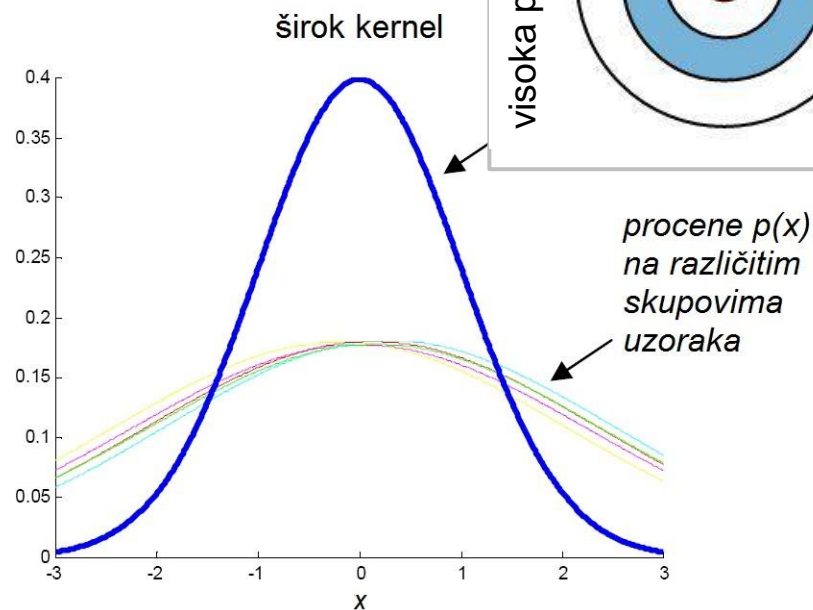
Uticaj izbora širine

Izbor širokog kernela

- smanjuje varijansu jer smanjuje uticaj pojedinih uzoraka
- povećava pristrasnost jer se procena p(x) bazira na većem području

Izbor uskog kernela

- smanjuje pristrasnost jer se procena p(x) bazira na manjem području
- povećava varijansu jer se procena p(x) bazira na većem području



Metode izbora širine kernela (1-D slučaj)

■ Subjektivni izbor

- Formirati procene za nekoliko vrednosti širine kernela i odabrati vrednost za koju se dobija procena koja najviše odgovara našim apriornim (subjektivnim) pretpostavkama
- Ovaj pristup je teško primenljiv u slučaju višedimenzionalnih obeležja

■ Za raspodele koje otprilike odgovaraju Gaussovima (unimodalne, simetrične, ne suviše izraženih „repova“)

- Bira se širina kernela koja minimizuje tzv. *integralnu srednjekvadratnu grešku* (MISE) u odnosu na Gaussovu raspodelu:

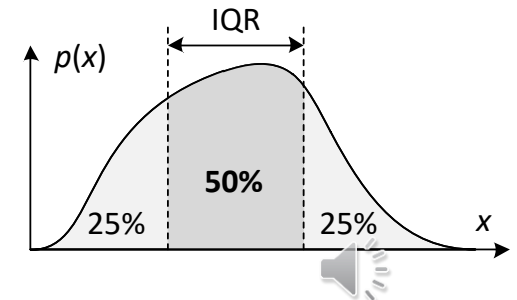
$$h_{OPT} = \underset{h}{\operatorname{argmin}} MISE\{\hat{p}_{KDE}(\mathbf{x})\} = \underset{h}{\operatorname{argmin}} E\left\{\int (\hat{p}_{KDE}(\mathbf{x}) - p(\mathbf{x}))^2 dx\right\}$$

- Za Gaussov kernel optimalna širina je $h_{OPT} = 1,06\sigma N^{-0,2}$ (Silvermanovo pravilo), gde je σ uzoračka standardna devijacija, a N broj uzoraka za obuku

■ Za multimodalne raspodele bolje je koristiti modifikovanu formulu, koja se zasniva na robustnijoj meri rasipanja:

$$h_{OPT} = 0,92AN^{-0,2}, \quad A = \min\left\{\sigma, \frac{IQR}{1,32}\right\}$$

gde je σ uzoračka standardna devijacija, a $IQR = Q3 - Q1$ tzv. *interkvartilni raspon* (razlika između 75. i 25. percentila GRV)



Metode izbora širine kernela (1-D slučaj)

- Direktna procena GRV na osnovu maksimalne izglednosti sama po sebi ne dolazi u obzir zato što bi dala $h_{ML} = 0$, odnosno, procenu GRV činili bi δ -impulsi u tačkama iz skupa za obuku
 - Međutim, moguća je unakrsna validacija izglednosti (eng. *likelihood cross-validation*), pri čemu se maksimizuje *pseudoizglednost*, što je ukupna izglednost svih n uzoraka, ali je za svaki uzorak računata u odnosu na raspodelu procenjenju na preostalih $N - 1$:

$$h_{MLCV} = \underset{h}{\operatorname{argmax}} \sum_{n=1}^N \ln p_{-n}(x^{(n)})$$
$$p_{-n}(x^{(n)}) = \frac{1}{(N-1)h} \sum_{\substack{m=1 \\ m \neq n}}^N K\left(\frac{x^{(n)} - x^{(m)}}{h}\right)$$

- Višedimenzionalni slučaj je znatno složeniji, pogotovo uzevši u obzir da su rasipanja po različitim dimenzijama u praksi različita, tako da širine kernela ne bi trebalo da budu iste po svim dimenzijama, tj. potreban je *vektor širine kernela*, a često i čitava *kovarijanska matrica*

Separabilni kerneli (eng. *product kernels*)

- Jedan vrlo popularan metod za KDE u višedimenzionalnom prostoru obuhvata upotrebu *separabilnih kernela*, koji se mogu predstaviti kao proizvod jednodimenzionalnih kernela:

$$\hat{p}_{PKDE}(\mathbf{x}) = \frac{1}{N} \cdot \frac{1}{h_1 h_2 \dots h_D} \sum_{n=1}^N \prod_{d=1}^D K_d \left(\frac{x_d - x_d^{(n)}}{h_d} \right)$$

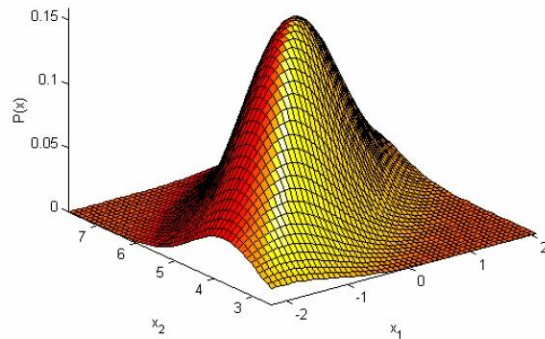
- Obično se ista funkcija $K_d(x)$ koristi za svaku dimenziju, a eventualno se razlikuju širine odgovarajućih 1-D kernela
- Određivanje širine kernela za svaku dimenziju može se izvršiti nekom od prethodno pomenutih metoda
- Separabilnost kernela ne implicira *nezavisnost pojedinačnih obeležja*
 - KDE metoda koja bi pretpostavljala nezavisnost obeležja imala bi sledeći oblik:

$$\hat{p}_{KDE_INDF}(\mathbf{x}) = \prod_{d=1}^D \frac{1}{N h_d} \sum_{n=1}^N K_d \left(\frac{x_d - x_d^{(n)}}{h_d} \right)$$

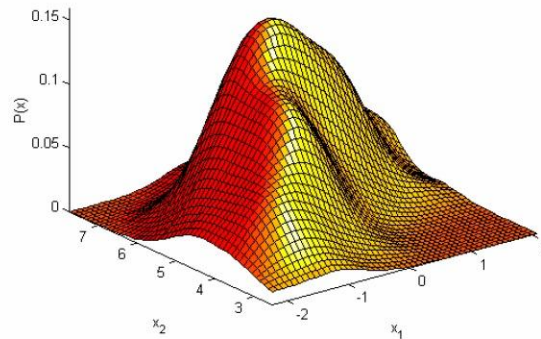
Separabilni kerneli (primer 1)

- Dvodimenzionalna unimodalna Gaussova raspodela, $N = 100$ uzoraka

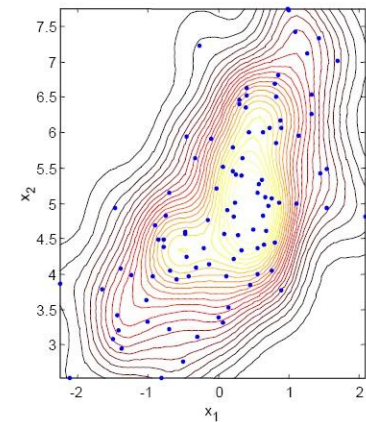
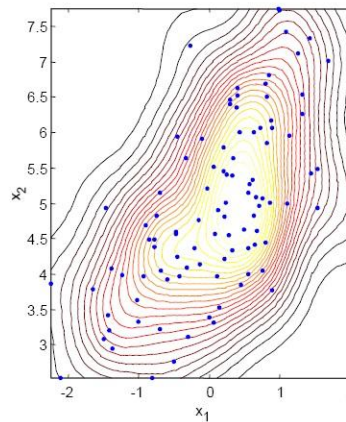
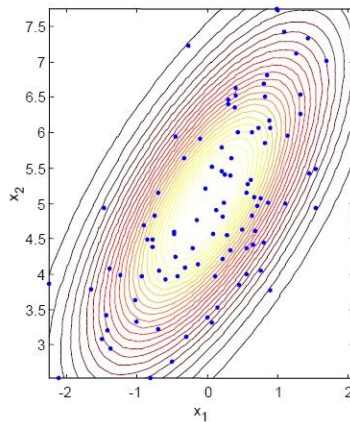
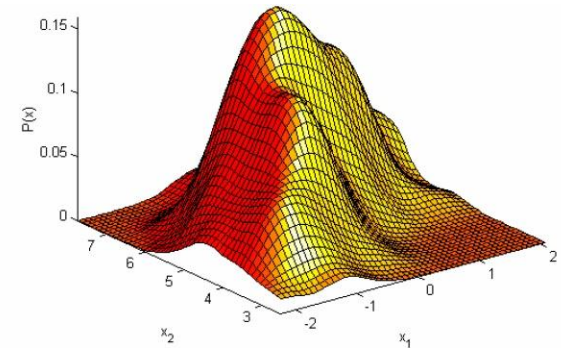
stvarna raspodela



PKDE, $h_{OPT} = 1,06\sigma N^{-0,2}$



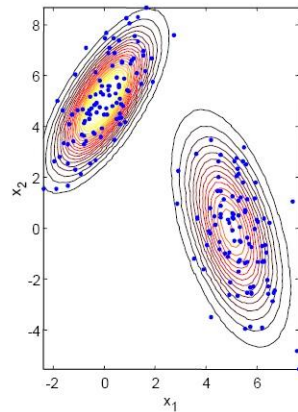
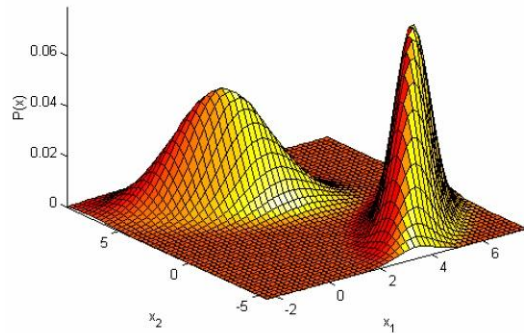
PKDE, $h_{OPT} = 0,92AN^{-0,2}$



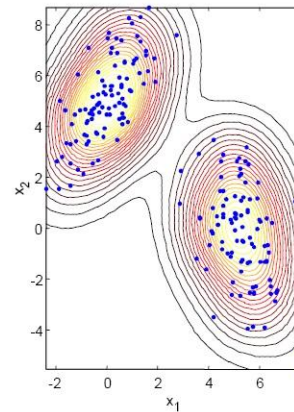
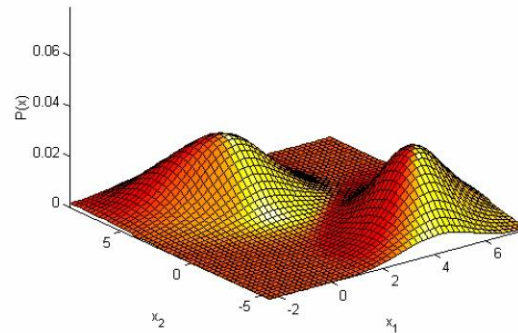
Separabilni kerneli (primer 2)

- Dvodimenzionalna bimodalna Gaussova raspodela, $N = 100$ uzoraka

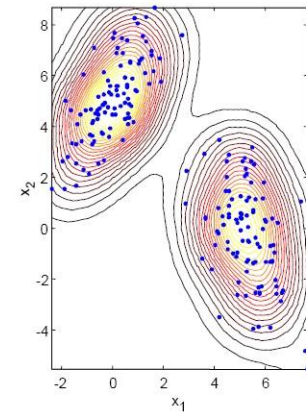
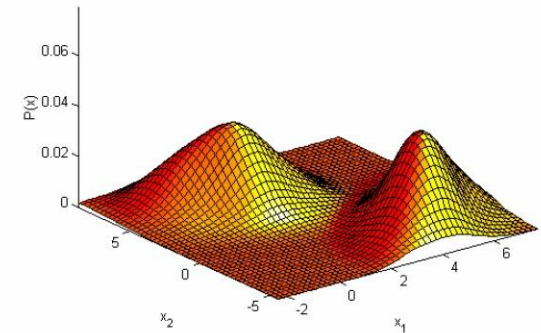
stvarna raspodela



PKDE, $h_{OPT} = 1,06\sigma N^{-0,2}$



PKDE, $h_{OPT} = 0,92AN^{-0,2}$



Naivni Bayesov klasifikator

- Bayesov klasifikator definisan je pravilom odlučivanja:

„dodeli \mathbf{x} klasi ω_i ako je $g_i(\mathbf{x}) \geq g_j(\mathbf{x})$ za svako $j \neq i$ “

pri čemu su diskriminantne funkcije jednake $g_i(\mathbf{x}) = P(\omega_i | \mathbf{x}) \propto p(\mathbf{x} | \omega_i)P(\omega_i)$

- $P(\omega_i)$ je u praksi definisano apriornim znanjem
 - $p(\mathbf{x} | \omega_i)$ se estimira npr. pomoću KDE, što u visokodimenzionalnom prostoru obeležja može biti velik problem
- Naivni Bayesov klasifikator predstavlja veoma praktično pojednostavljenje Bayesovog klasifikatora
 - Pretpostavka je da su obeležja nezavisna u okviru iste klase:

$$p(\mathbf{x} | \omega_i) = \prod_{d=1}^D p_d(x_d | \omega_i)$$

što je mnogo manje striktno nego pretpostavka da su generalno nezavisna: $p(\mathbf{x}) = \prod_{d=1}^D p_d(x_d)$

- Glavna prednost ove metode leži u tome što je umesto višedimenzionalnih GRV $p(\mathbf{x} | \omega_i)$ dovoljno estimirati samo jednodimenzionalne GRV $p_d(x_d | \omega_i)$
- Uprkos jednostavnosti, naivni Bayesov klasifikator ima veoma dobre performanse

Naivni Bayesov klasifikator (primer)

- Neka se razmatra Bayesov klasifikator čiji je zadatak klasifikacija D -dimenzionalnih vektora obeležja $\mathbf{x} = [x_1, \dots, x_D]^T$ sa binarnim vrednostima ($x_i \in \{0, 1\}$, $i = 1, \dots, D$), i neka su verovatnoće da i -ta koordinata vektora \mathbf{x} bude jednaka 1 u pojedinim klasama jednake $P(x_i = 1 | \omega_1) = p_i$ i $P(x_i = 1 | \omega_2) = q_i$.
- Po Bayesovom pravilu, dati vektor \mathbf{x} klasifikuje se na osnovu odnosa izglednosti:

$$\Lambda(\mathbf{x}) = \frac{P(\mathbf{x} | \omega_1)}{P(\mathbf{x} | \omega_2)} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{P(\omega_2)}{P(\omega_1)}$$

- Međutim, bez pretpostavke o nezavisnosti u okviru pojedinih klasa, nastaje problem sa količinom potrebnih uzoraka za estimaciju $P(\mathbf{x} | \omega_i)$ jer \mathbf{x} može imati 2^D mogućih vrednosti, pa je za svaku klasu potrebno po $2^D - 1$ pouzdanih estimacija
- Uz pretpostavku o nezavisnosti obeležja u okviru pojedinih klasa, odgovarajuće verovatnoće jednake su:

$$P(\mathbf{x} | \omega_1) = \prod_{i=1}^D p_i^{x_i} (1 - p_i)^{1-x_i}$$
$$P(\mathbf{x} | \omega_2) = \prod_{i=1}^D q_i^{x_i} (1 - q_i)^{1-x_i}$$

i broj potrebnih estimacija po klasi sada je D (umesto $2^D - 1$)

$D = 3$ (klasa ω_1)

Mogući ishodi (\mathbf{x})	$P(\mathbf{x} \omega_1)$
000	0.10
001	0.15
010	0.10
001	0.25
100	0.05
101	0.05
110	0.20
111	0.10