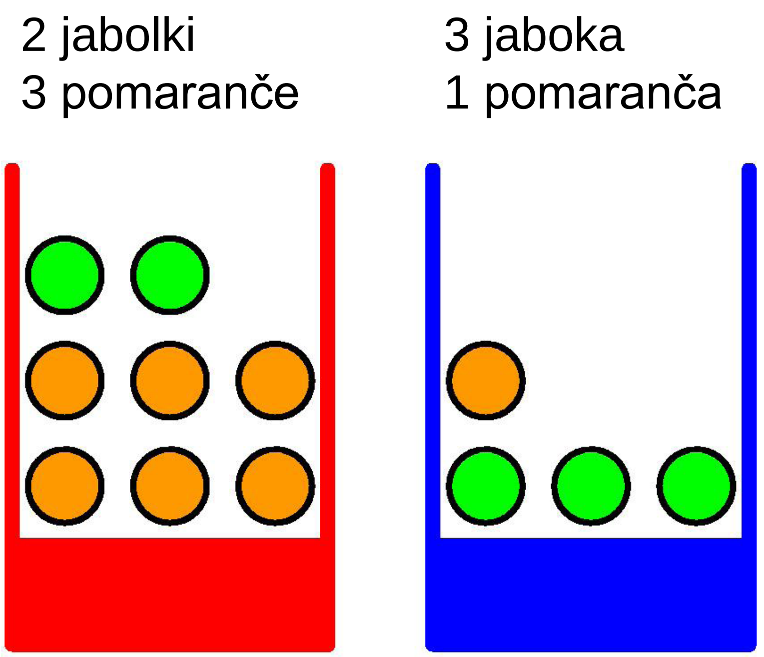
LABORATORIJ ZA DIGITALNO OBDELAVO SIGNALOV, SLIK IN VIDEA
FAKULTETA ZA ELEKTROTEHNIKO, UNIVERZA V LJUBLJANI

Teorija verjetnosti
uvod

prof. dr. Jurij Tasič
Asistent Emil Plesnik
Laboratorij za digitalno obdelavo signalov, slik in videa
http://www.ldos.si/

Teorija verjetnosti z več spremeljivkami

- Ključni koncept je delovanje z negotovostjo – zaradi šuma in končnih nizov podatkov
- Osnova za kvantizacijo in upravljanje z negotovostjo
- Vzemimo, da vlečemo iz rdeče škatle 40% potez iz modre pa 60%



Škatla je naključna spremenljivka B (zavzema vrednosti *r ali b)*
Sadje je naključna spremenljivka F (zavzema vrednosti *o ali a)*
Naj bo $p(B=r)=4/10$ *in* $p(B=b)=6/10$

Osnovne informacije /1

- Marginalna verjetnost

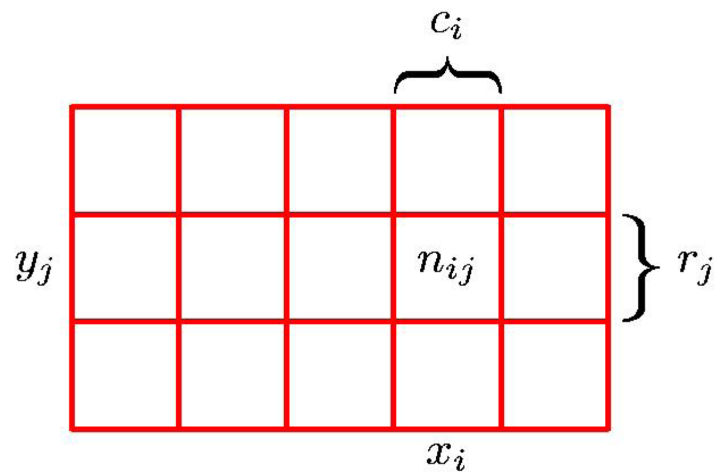
– kakšna je verjetnost jabolka?

- Pogojna verjetnost

– Pod pogojem, da imamo pomarančo, kakšna je verjetnost, da bomo izbrali modro škatlo? (0.6)

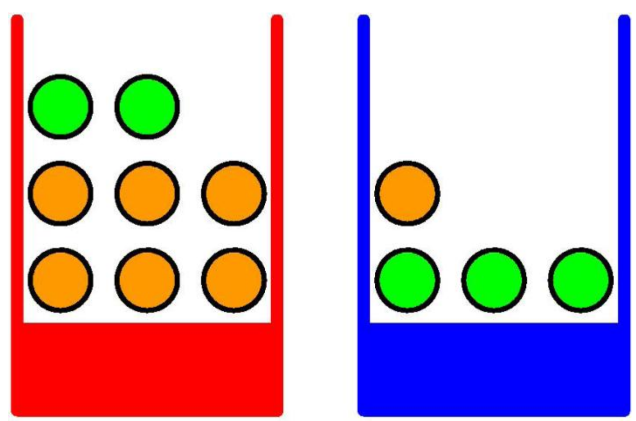
- Skupna verjetnost

– Kakšna je verjetnost pomaranče IN modre škatle?



2 jabolki
3 pomaranče

3 jaboka
1 pomaranča



Joint Probability $p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$

$$p(X = x_i) = \frac{c_i}{N}$$

Since $c_i = \sum_j n_{ij}$, $p(X = x_i) = \sum_{j=1}^{L} p(X = x_i, Y = y_j)$

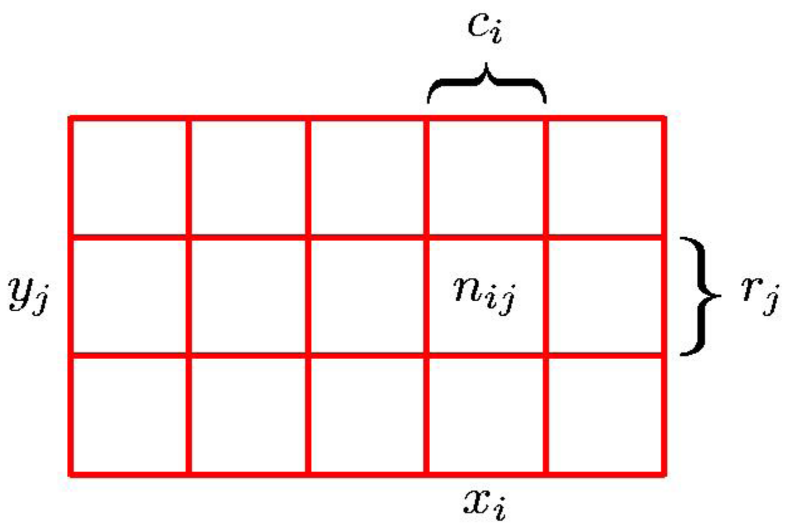
Pravilo vsote teorije verjetnosti

- Imamo dve naključni spremenljivki
- *X lahko zavzame vrednosti*

  $x_i, \quad i=1,..,M$
- *Y lahko zavzame vrednosti*

  $y_i, \quad i=1,..,L$
- *N je število poskusov za X in Y*
- Število poskusov, kjer je

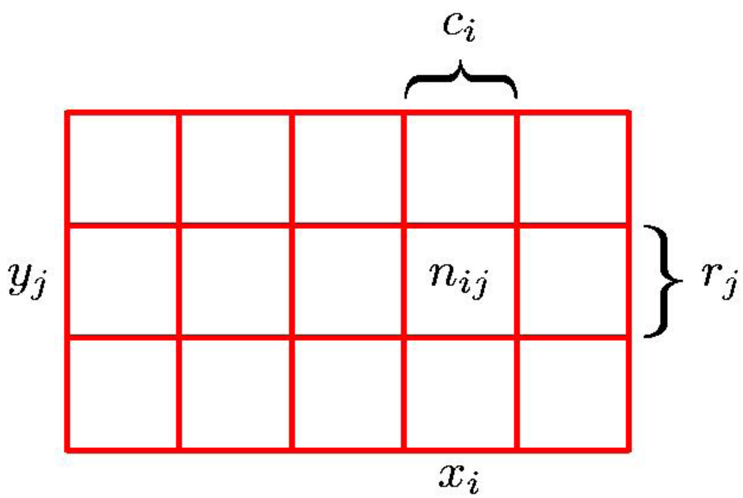
  $X=x_i$ ter $Y=y_j$ je n_{ij}



Marginalna verjetnost

Pravilo produkta teorije verjetnosti

- Obravnavajmo zgolj primere, kjer je $X=x_i$
- Potem lahko del primerov, kjer je $Y=y_j$ zapišemo kot $p(Y=y_j/X=x_i)$
- To je pogojna verjetnost
- Relacija med skupno in pogojno verjetnostjo:



$$p(Y=y_j \mid X=x_i)=\frac{n_{ij}}{c_i}$$

$$p(X=x_i,Y=y_j)=\frac{n_{ij}}{N}=\frac{n_{ij}}{ci}\bullet\frac{c_i}{N}$$

$$=p(Y=y_j \mid X=x_i)p(X=x_i)$$

Bayesov teorem

- Z upoštevanjem pravila produkta in lastnosti simetričnosti $p(X,Y)=p(Y,X)$ dobimo

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}.$$

- Temu pravimo Bayesov teorem
- Glede na pravilo produkta, lahko izrazimo imenovalec

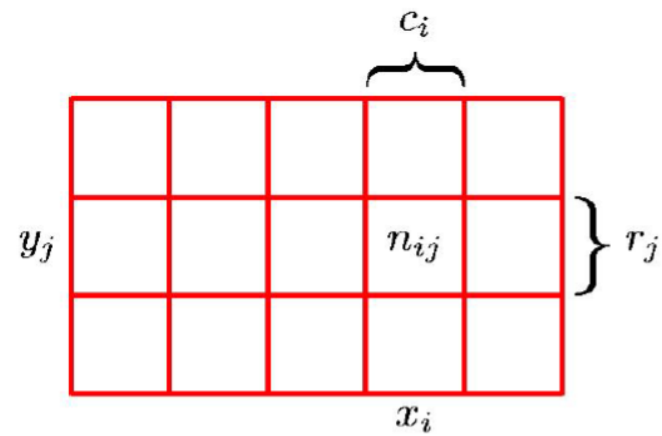
$$p(X) = \sum_Y p(X|Y)p(Y)$$

Normalizacijska konstanta zagotavlja, da je vsota pogojne verjetnosti za LHS enaka 1 za vse vrednosti Y

Pravila verjetnosti

- Imamo naključni spremenljivki X in Y
- Pravilo vsote daje marginalno verjetnost

$$p(X = x_i) = \sum_{j=1}^{L} p(X = x_i, Y = y_j) = \frac{c_i}{N}$$



- Pravilo produkta: skupna verjetnost glede na pogojno in marginalno

$$p(X,Y) = \frac{n_{ij}}{N} = p(Y \mid X)p(X) = \frac{n_{ij}}{c_i} \times \frac{c_i}{N}$$

- S kombiniranjem dobimo Bayesov teroem

$$p(Y \mid X) = \frac{p(X \mid Y)p(Y)}{p(X)}$$ kjer je $$p(X) = \sum_{Y} p(X \mid Y)p(Y)$$

To je videno kot Posteriorna α verjetnost X

Skupna verjetnost dveh spremenljivk

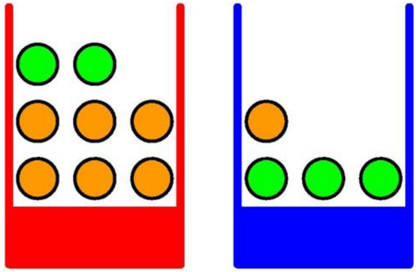
Vrnimo se k primeru škatel in sadja. Verjetnost izbire rdeče ali modre škatle je podana kot

$$p(B = r) = 4/10 \qquad p(B = b) = 6/10$$

Velja $\quad p(B = r) + p(B = b) = 1.$

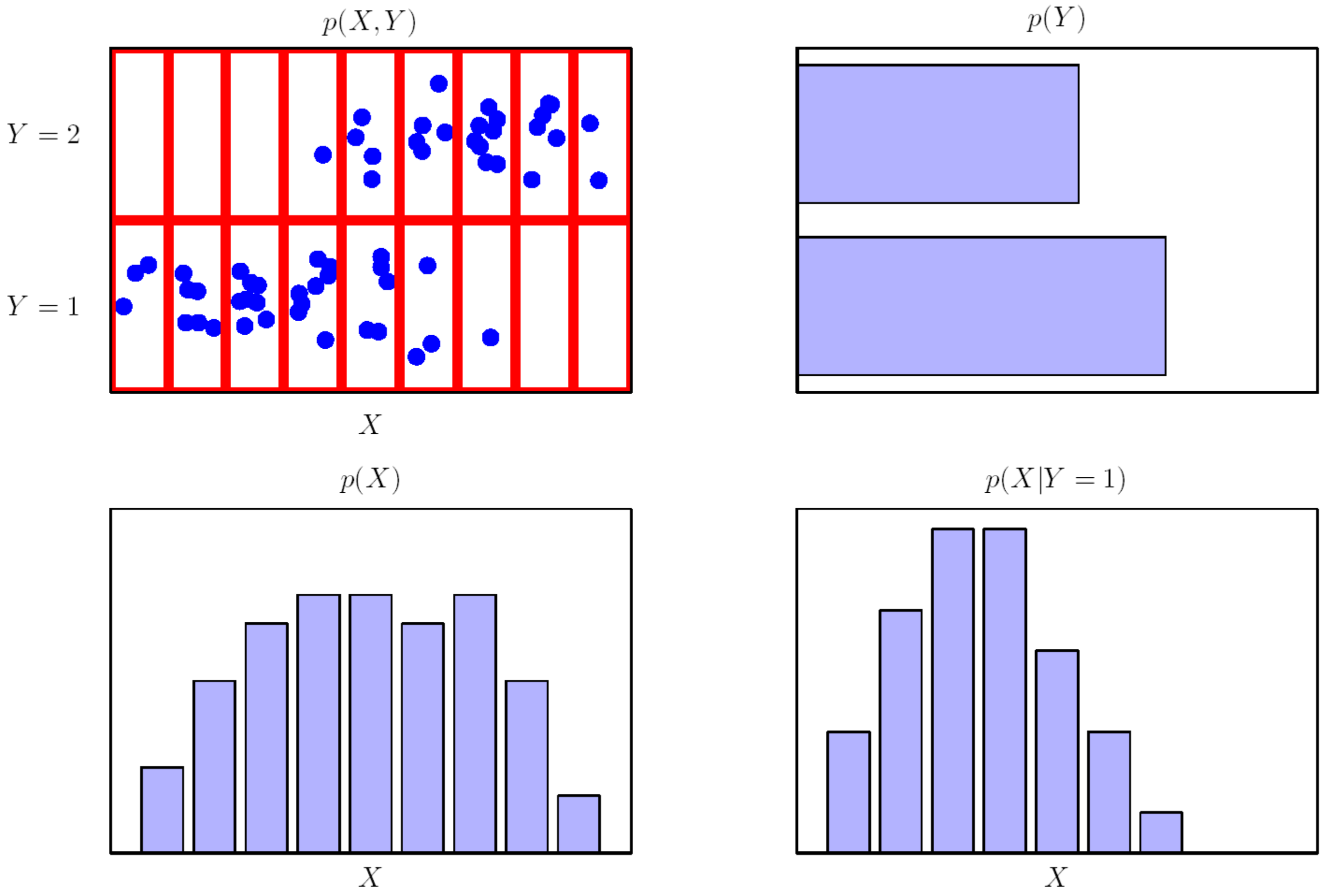
Naključno izberemo modro škatlo. Potem je verjetnost izbire jabolka odvisna zgolj od deleža jabolk v modri škatli, ki je 3/4 in je torej
$p(F = a|B = b) = 3/4.$
Zapišemo lahko pogojne verjetnosti za obe vrsti sadja in škatel

$$p(F = a|B = r) = 1/4$$
$$p(F = o|B = r) = 3/4$$
$$p(F = a|B = b) = 3/4$$
$$p(F = o|B = b) = 1/4.$$



Skupna verjetnost dveh spremenljivk

X zavzema 9 vrednosti, Y le 2



Slika zgoraj levo prikazuje 60 točk kot rezultat skupne verjetnosti teh dveh spremenljivk. Ostale slike prikazujejo histograme marginalnih porazdelitev p(X) in p(Y), kot tudi pogojno porazdelitev p(X|Y = 1).

Uporaba Bayesovega teorema na primeru s sadjem

- Verjetnosti so normalizirane, da velja

$$p(F = a|B = r) + p(F = o|B = r) = 1$$

In podobno za

$$p(F = a|B = b) + p(F = o|B = b) = 1.$$

- Skupno verjetnost izbire jabolka lahko ocenimo z uporabo pravil vsote in produkta

$$\begin{aligned} p(F = a) &= p(F = a|B = r)p(B = r) + p(F = a|B = b)p(B = b) \\ &= \frac{1}{4} \times \frac{4}{10} + \frac{3}{4} \times \frac{6}{10} = \frac{11}{20} \end{aligned}$$

iz česar z uporabo pravila vsote sledi

$$p(F = o) = 1 - 11/20 = 9/20$$

Skupna verjetnost dveh spremenljivk

Vemo, da je bila v poizkusu izbrana pomaranča. Sedaj nas zanima, iz katere škatle je ta pomaranča. Zato moramo določiti verjetnostno porazdelitev škatel glede na vrsto sadja. Pri tem nam verjetnosti dajejo verjetnostno porazdelitev sadja glede vrsto škatle.

Problem lahko rešimo tako, da z uporabo Bayesovega teorema obrnemo pogojno verjetnost:

$$p(B=r|F=o) = \frac{p(F=o|B=r)p(B=r)}{p(F=o)} = \frac{3}{4} \times \frac{4}{10} \times \frac{20}{9} = \frac{2}{3}$$

Iz pravila vsote potem sledi $p(B=b|F=o) = 1 - 2/3 = 1/3$

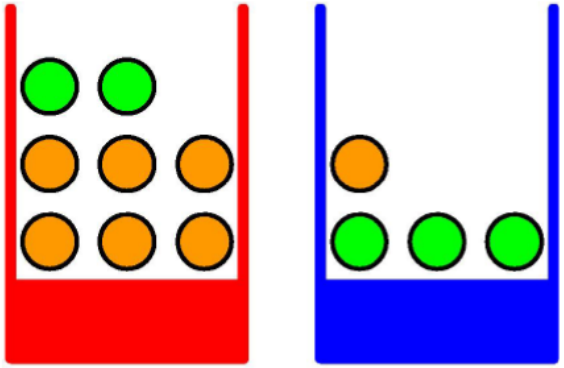
Interpretacija Bayesovega teorem je laho tudi sledeča. Če bi se pojavilo vprašanje izbire škatle preden smo izvedeli za izbiro sadja, potem bi bila najboljša možna informacija na voljo z verjetnostjo p(B). Temu pravimo apriori verjetnost, saj je verjetnost na voljo preden izvemo za izbrano sadje. Ko pa enkrat že vemo, da je izbrana pomaranča, lahko uporabimo Bayesov teorem, da določimo verjetnost p(B|F), ki ji pravimo aposteriori verjetnost, ker je pridobljena po znanem rezultatu poskusa F.

Uporaba Bayesovega teorema na primeru s sadjem

- Verjetnost, da je škatla rdeča ob pogoju, da je izbrana pomaranča je

$$p(B=r \mid F=o) = \frac{p(F=o \mid B=r)\,p(B=r)}{p(F=o)}$$

$$= \frac{\frac{3}{4} \times \frac{4}{10}}{\frac{9}{20}} = \frac{2}{3} = 0.66$$



A posteriori verjetnost 0.66 je različna od *a priori* verjetnosti 0.4

- Verjetnost, da je izbrana pomaranča
  - glede na pravili vsote in produkta

Marginala verjetnost 0.45 je manjša *od povprečne* verjetnosti *7/12=0.58*

$$p(F=o) = p(F=o, B=r) + p(F=o, B=b)$$

$$= p(F=o \mid B=r)\,p(B=r) + p(F=o \mid B=b)\,p(B=b)$$

$$= \frac{6}{8} \times \frac{4}{10} + \frac{1}{4} \times \frac{6}{10} = \frac{9}{20} = 0.45$$

Neodvisne spremenljivke

- Če je $p(X,Y) = p(X)p(Y)$, potem sta X in$Y$ neodvisni spremenljivki
- Zakaj?
- Iz pravila produkta sledi $p(Y \mid X) = \dfrac{p(X,Y)}{p(X)} = p(Y)$

- Če v primeru škatel in sadja vsaka škatla vsebuje enak delež jabolk in pomaranč, potem velja

$$p(F|B)=p(F)$$

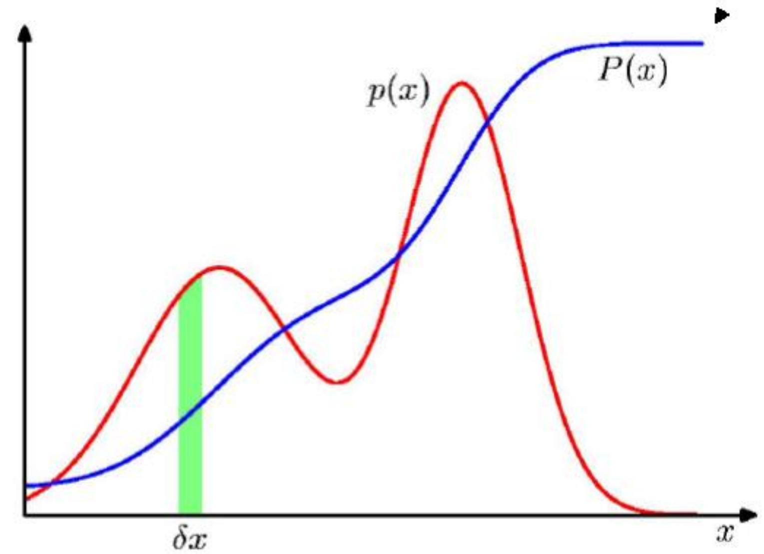
Gostota verjetnosti

- Zvezne spremenljivke
- Če je verjetnost, da je x na intervalu $(x, x + \delta x)$ podana s $p(x)dx$ za δx $--> 0$, *potem* je $p(x)$ pdf funkcija x

- Verjetnost, da x leži na intervalu (a, b) je

$$p(x \in (a,b)) = \int_a^b p(x)dx$$

Kumulativna porazdelitvena funkcija



Verjetnost, da x leži na intervalu $(-\infty, z)$ je

$$P(z) = \int_{-\infty}^{z} p(x)dx$$

Več spremenljivk

- Če imamo več zveznih spremenljivk
  $x1, \ldots, xD$ označenih z vektorjem x, potem lahko določimo skupno gostoto verjetnosti

$$p(x) = p(x1, \ldots, xD)$$

- Gostota verjetnosti več spremenljivk mora zadostiti pogojema

$$p(x) \geq 0$$

$$\int p(x) dx = 1$$

Vsota, produkt in Bayes v zveznem postoru

- Pravila veljajo tudi v primerih zveznih in kombiniranih diskretnih in zveznih spremenljivk

$$p(x) = \int p(x,y)dy$$

$$p(x,y) = p(y \mid x)p(x)$$

$$p(y \mid x) = \frac{p(x \mid y)p(y)}{p(x)}$$

Matematično upanje(Expectation)

- Matematično pričakovanje je povprečje neke funkcije f(x) glede na porazdelitev verjetnosti p(x) in jo označimo kot E[f]
- Za diskretno porazdelitev velja

$$E[f] = \sum_x f(x)p(x)$$

- Za zvezno porazdelitev velja

$$E[f] = \int f(x)p(x)dx$$

- Če imamo N točk iz pdf, potem lahko pričakovanje ocenimo kot

$$E[f] = \frac{1}{N}\sum_1^N f(x_n)$$

- Pogojno pričakovanje glede na pogojno porazdelitev je

$$E_x[f] = \sum_x p(x|y)f(x)$$

Varianca

- Podaja variabilnost v $f(x)$ okrog njene srednje vrednosti $E[f(x)]$
- Varianca $f(x)$ je označena kot

$$var[f] = E[(f(x) - E[f(x)])^2]$$

- *Če razširimo kvadrat*

$$var[f] = E[(f(x)^2] - E[f(x)]^2$$

- Varianca spremenljivke x

$$var[x] = E[x^2] - E[x]^2$$

Kovarianca

- Za dve naključni spremenljivki x in y je kovarianca določena kot

$$cov[x,y] = E_{x,y}[\{x-E[x]\}\{y-E[y]\}]$$
$$= E_{x,y}[xy] - E[x]E[y]$$

- Izraža, kako x in y variirata skupaj
- Če sta x in y neodvisna, potem med njima ni kovariance
- Če sta x in y dva vektorja naključnih spremenljivk je kovarianca v obliki matrike
- Če določamo kovarianco med komponentami vektorja x, jo označimo kot

$$cov[x] = cov[x,x]$$

Bayesova verjetnost

- Klasični ali Frekvenčni vidik verjetnosti
  - Verjetnost je frekvenca naključnega, ponovljivega dogodka
  - Frekvenca simbola pri metu kovanca je 1/2
- Bayesov vidik
  - Verjetnost je kvantifikacija negotovosti
  - Stopnja verjetja v predloge, ki ne vključujejo naključnih spremenljivk
  - Primeri negotovih dogodkov kot verjetnosti:
    - Ali je Shakespeareova dela napisal Francis Bacon
    - Ali je Luna nekoč krožila v svoji lastni orbiti okrog Sonca
    - Ali je podpis na čeku pristen

Primeri negotovih dogodkov

- Ali bo arktični ledeni pokrov izginil do konca stoletja
  - Imamo nekaj podatkov o hitrosti taljenja polarnega ledu
  - Preučimo jih ponovno na podlagi novih dokazov, podatkov (satelitske slike)
  - Ocena bo vplivala na prihodnja dejanja (zmanjševanje toplogrednih plinov)
- Vse to lahko dosežemo s splošno Bayesovo interpretacijo

Bayesova predstavitev negotovosti

- Predstavitev negotovosti z verjetnostjo ni ’’ad-hoc’’ izbira
- Če za predstavitev stopnje verjetja uporabimo numerične vrednosti, potem preprosti aksiomi za spreminjanje stopnje verjetja vodijo k pravilom vsote in produkta za verjetnost (Coxov teorem)
- Teorijo verjetnosti lahko obravnavamo kot nagradnjo Boolove logike za primere, ki vključujejo negotovost (Jaynes)

Bayesov pristop

- Kvantiziramo negotovost okrog izbranih parametrov **w**
  - npr., w je vektor parametrov pri prilagajanju krivulj
- Negotovost pred pregledom podatkov izrazimo kot $p(\mathbf{w})$
- Glede na podatke $D = \{ t_1, . . t_N \}$
  - negotovost v w po pregledu D, po Bayesovem pravilu:

$$p(\mathrm{w} \,|D) = p(D\,|\mathrm{w})p(\mathrm{w}) \,/\, p(D)$$

- Kvantiteta $p(D|\mathbf{w})$ je lahko tudi funkcija vektorja w
  - kako verjeten je niz podatkov za različne parametre w
    - Imenuje se funkcija verjetja
    - Ni verjetnostna porazdelitev parametrov w

Vloga funkcije verjetja

- Negotovost v w izrazimo:

$$p(\mathrm{w}\,|D) = \frac{p(D\,|\mathrm{w})p(\mathrm{w})}{p(D)}$$

  Kjer je $p(D) = \int p(D\,|\mathrm{w})p(\mathrm{w})d\mathrm{w}$ po pravilu vsote
- Imenovalec je normalizacijski faktor
  - Vključuje marginalizacijo po w
- Bayesov teorem v besedah

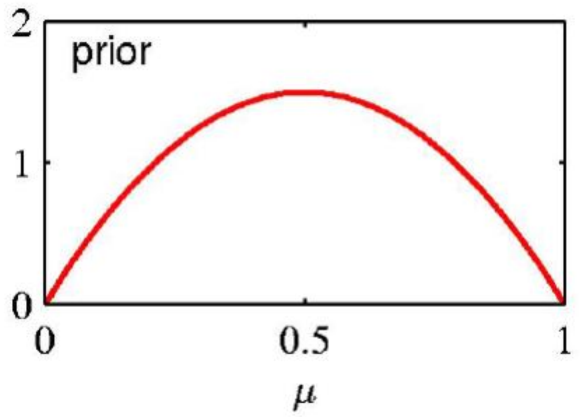
  Posteriorna α verjetnost x prirona
- Verjetnostna funkcija igra osrednjo vlogo tako v Bayesovem kot ciljnem pristopu
- Ciljni pristop: w je fiksni parameter določen z oceno; približna napaka več možnih podatkovnih naborov D
- Bayesov pristop: imamo en sam podatkovni nabor D, negotovost je izražena kot verjetnostna porazdelitev po w

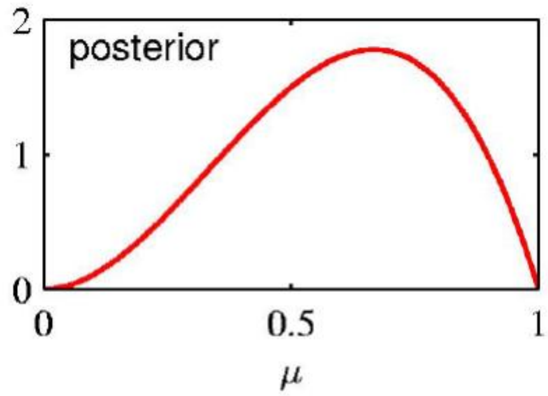
Pristop največjega verjetja

- Po Frekventistni postavitvi je w fiksni parameter
  - w je postavljen na vrednost, ki maksimizira funkcijo verjetja *p(D*|w*)*
  - Pri največjem verjetju je napaka negativni log funkcije verjetja, saj največje verjetje pomeni minimizacijo napake
  - Bootstrap pristop k ustvarjanju *L* zbirk podatkov
    - Iz N podatkovnih točk lahko ustvarimo nove zbirke podatkov z naključno zamenjavo N točk.
    - Z L ponovitvami ustvarimo L zbirk podatkov
    - Natančnost približka parametrov lahko ocenimo z variabilnostjo predikcij med različnimi bootstrap zbirkami

Bayes vs Frequentist

- Vključitev predhodnega znanja se pojavi naravno
- Primer meta kovanca
  - Kovanec vržemo trikrat in vedno pristane na simbolu
  - Klasični pristop največjega verjetja za verjetnost simbola nakazuje, da bo rezultat vseh prihodnjih metov simbol
  - Bayesov pristop z upoštevanjem predhodnjega znanja vodi do manj ekstremnih zaključkov





Praktičnost Bayesovega pristopa

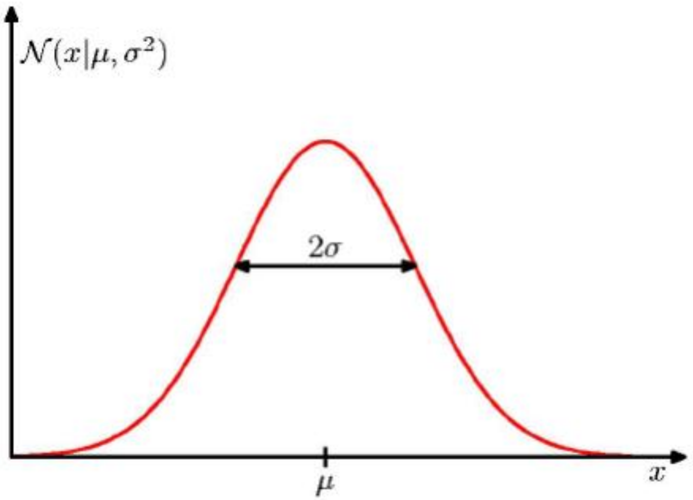
- Marginalizacija prek celotnega prostora parametrov je potrebna za predikcijo ali primerjavo modelov
- Praktični faktorji:
  - Vzorčevalne metode, npr. Markov Chain Monte Carlo metode
  - Povečana hitrost in količina spomina pri računalnikih
- Deterministična aproksimacija kot npr. variacijski Bayes in širjenje pričakovanja (ang. Expectation propagation) so alternative vzorčenju

Gaussova porazdelitev

- Za eno samo realno spremenljivko
- Parametri:

$$N(x \mid \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

  – Srednja vrednost μ, varianca σ^2
- *Standardna deviacija* σ^2
- *Preciznost* $\beta = 1/\sigma^2$



- $E[x] = \mu$
- $Var[x] = \sigma^2$

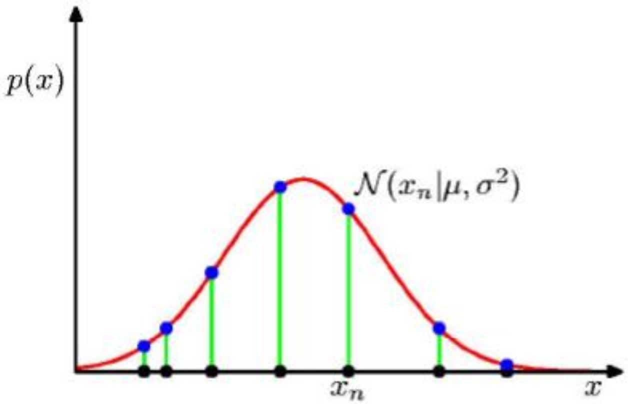
Maksimum porazdelitve pomeni njen način

Za Gaussovo porazdelitev način sovpada z njeno srednjo vrednostjo

Funkcija verjetja za Gaussovo p.

- Imamo *N* podatkov x_i, $i=1,..,n$
- Neodvisni in enakomerno porazdeljeni
- Verjetnost zbirke podatkov je podana s funkcijo verjetja

$$p(x \mid \mu, \sigma^2) = \prod_{n=1}^{N} N(x_n \mid \mu, \sigma^2)$$



- Logaritemska funkcija verjetja je

$$\ln p(x \mid \mu, \sigma^2) = -\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2 - \frac{N}{2}\ln\sigma^2 - \frac{N}{2}\ln(2\pi)$$

- Rešitve največjega verjetja so

$$\mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} x_n$$

$$\sigma_{ML}^2 = \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu_{ML})^2$$

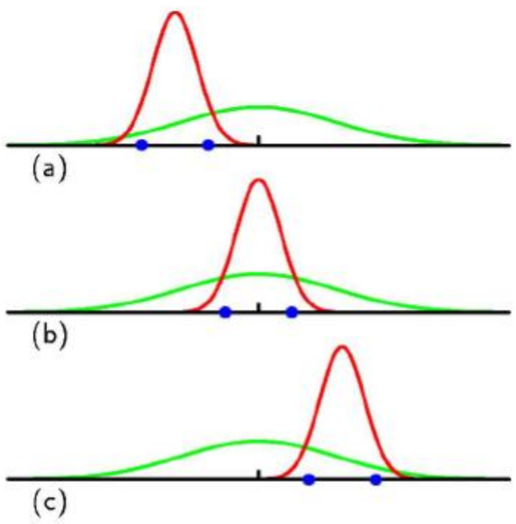
Podatki: črne točke

Verjetje= produkt modrih vrednosti

Izberemo srednjo vrednost in varianco in maksimiziramo produkt

Pristranskost v največjem verjetju

- Največje verjetje sistematično zapostavlja varianco
  - $E[\mu_{ML}]=\mu$
  - $E[\sigma^2_{ML}]=((N-1)/N)\sigma^2$



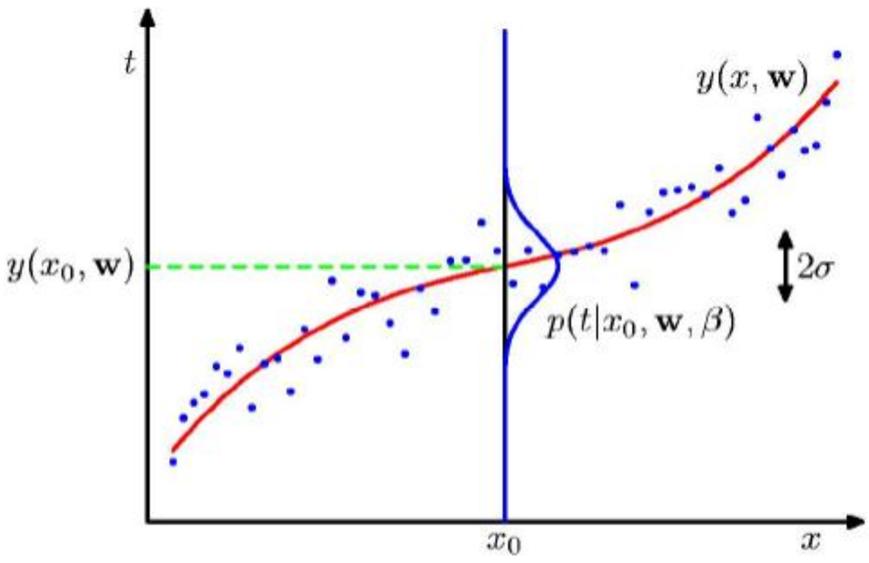
Problem je podoben ”overfitting” problemu.

Srednja vrednost določena prek treh zbirk podatkov je pravilna.
Varianca je zapostavljena, ker je ocenjena relativno glede na srednjo vrednost vzorca in ne na pravo srednjo vrednost.

Verjetnostno prilagajanje krivulj

- Cilj je predvideti vrednosti ciljne spremenljivke *t* glede na novo vrednost vhodne spremenljivke *x*
- Imamo *N* vhodnih vrednosti $x=(x1,..xN)^T$ in pripadajoče ciljne vrednosti $t=(t1,..,tN)^T$
- Predvidevamo, da ima glede na dano vrednost *x*, vrednost *t* Gaussovo porazdelitev s srednjo vrednostjo *y(x,w)* polinomske krivulje

$$p(t|x,\mathrm{w},\beta)=N(t|y(x,\mathrm{w}),\beta^{-1})$$



Gaussova pogojna porazdelitev za t za dani x.
Srednja vrednost je podana s polinomsko funkcijo *y(x,w)*
Preciznost je podana z β

Prilagajanje krivulj z največjo stopnjo verjetja

- Funkcija verjetja je $p(\mathrm{t} \mid \mathrm{x}, \mathrm{w}, \beta) = \prod_{n=1}^{N} N(t_n \mid y(x_n, \mathrm{w}), \beta^{-1})$
- Logaritem funkcije verjetja je

$$\ln p(t \mid x, \mathrm{w}, \beta) = -\frac{\beta}{2} \sum_{n=1}^{N} \{y(x_n, w) - t_n\}^2 + \frac{N}{2} \ln \beta - \frac{N}{2} \ln(2\pi)$$

- Rešitev največje stopnje verjetja za polinomske koeficiente w_{ML}
  - Maksimizacija w.r.t w
  - Opustimo lahko zadnja dva člena – nista odvisna od w
  - $\beta/2$ lahko zamenjamo z ½
  - Minimizacija negativnega log-verjetja
  - identično funkciji napake vsote kvadratov

Parameter preciznosti največjega verjetja

- Stopnja največjega verjetja je lahko uporabljena za določitev β Gaussove pogojne porazdelitve
- Rezultat maksimizacije verjetja wrt β

$$\frac{1}{\beta_{ML}} = \frac{1}{N}\sum_{n=1}^{N}\left\{y(x_n, w_{ML}) - t_n\right\}^2$$

- Najprej določimo vektor parametrov w_{ML},

  ki določa srednjo vrednost in posledično uporabo β

Prediktivna porazdelitev

- Poznavanje parametrov w in β
- Predikcije za nove vrednosti *x* lahko določimo z uporabo

$$p(t|x, \mathrm{w}_{ML}, \beta_{ML}) = N(t|y(x, \mathrm{w}_{ML}), \beta_{ML}{}^{-1})$$

- Namesto približkov točk sedaj podajamo verjetnostno porazdelitev nad *t*

Bayesov pristop

- Vpeljemo predhodno porazdelitev nad polinomom

$$p(\mathbf{w} \mid \alpha) = N(\mathbf{w} \mid 0, \alpha^{-1} I) = \left(\frac{\alpha}{2\pi} \right)^{(M+1)/2} \exp\left\{ -\frac{\alpha}{2} \mathbf{w}^T \mathbf{w} \right\}$$

kjer je α preciznost -natančnost porazdelitve

$M+1$ je skupno število parametrov polinoma Mte stopnje

A posteriori porazdelitev

- Z uporabo Bayesovega teorema je a posteriori porazdelitev za w proporcionalna produktu a priori porazdelitve in funkcije verjetja

$$p(\mathrm{w}|\mathrm{x},\mathrm{t},\alpha,\beta) \quad \propto \quad p(\mathrm{t}|\mathrm{x},\mathrm{w},\beta)p(\mathrm{w}|\alpha)$$

- W lahko določimo z najbolj verjetno vrednostjo w glede na podatke, to je z maksimizacijo aposteriori porazdelitve
- To je ekvivalentno minimizaciji (uporaba log)

$$\frac{\beta}{2}\sum_{n=1}^{N}\left\{y(x_n,\mathrm{w})-t_n\right\}^2+\frac{\alpha}{2}\mathrm{w}^T\mathrm{w}$$

- Enako kot funkciji vsote kvadratov napake z regulacijskim parametrom $\lambda=\alpha/\beta$

Bayesovo prilagajanje krivulj

- Predhodne metode še operirajo s točkovnimi približki vektorja w
  - v polnem Bayesovem pristopu uporabimo pravila vsote in produkta in integracijo čez vse vrednosti w
- Glede na podatke x in t in novo testno točko x, je naš cilj napovedati vrednost t
  - *t.j., oceniti prediktivno porazdelitev* $p(t|x,\mathrm{x},\mathrm{t})$
- Uporaba pravil vsote in produkta

$$p(t \mid x,\mathrm{x},\mathrm{t}) = \int p(t,\mathrm{w} \mid x,\mathrm{x},t)\,d\mathrm{w} \qquad \text{by Sum Rule (marginalizing over w)}$$

$$= \int p(t \mid x,\mathrm{w},\mathrm{x},\mathrm{t})\, p(\mathrm{w} \mid x,\mathrm{x},\mathrm{t}) \qquad \text{by Product Rule}$$

$$= \int p(t \mid x,\mathrm{w})\, p(\mathrm{w} \mid \mathrm{x},\mathrm{t})\,d\mathrm{w} \qquad \text{by eliminating unnecessary variables}$$

$$p(t \mid x,\mathrm{w}) = N(t \mid y(x\ ,\mathrm{w}),\beta^{-1})$$

Bayesovo prilagajanje krivulj

- A posteriori je lahko Gaussova

$$p(t \mid x, \mathbf{x}, \mathbf{t}) = N(t \mid m(x), s^2(x))$$

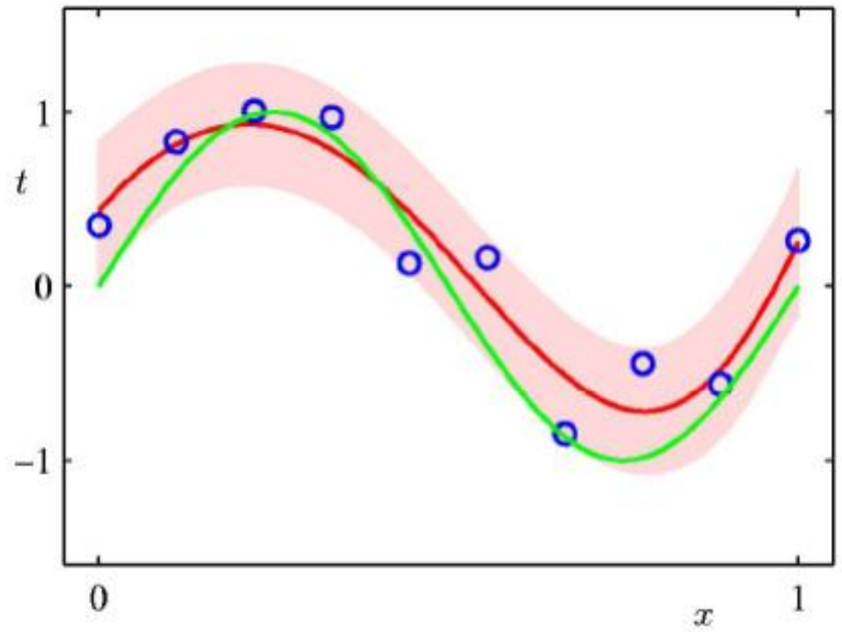
- Srednja vrednost in varianca sta odvisni od x

$$m(x) = \beta \phi(x)^T S \sum_{n=1}^{N} \phi(x_n) t_n$$

$$s^2(x) = \beta^{-1} + \phi(x)^T S \phi(x)$$

$$S^{-1} = \alpha I + \beta \sum_{n=1}^{N} \phi(x_n) \phi(x)^T$$

$\phi(x)$ has elements $\phi_i(x) = x^i$ for $i = 0,..M$



Prediktivna porazdelitev
M=9 polinom
$\alpha = 5 \times 10^{-3}$
$\beta = 11.1$
Rdeča krivulja je srednja vrednost
Rdeče območje je +1 std dev

Izbira modela

Modeli pri prilagajnju krivulj

Polinosko prilagajanje:
– optimalna stpnja polinoma daje najboljše rezultate

• Stopnja polinoma vpliva na število prostih parametrov v modelu in torej določa kompleksnost modela
• pri najmanjših kvadratih λ prav tako določa kompleksnost modela

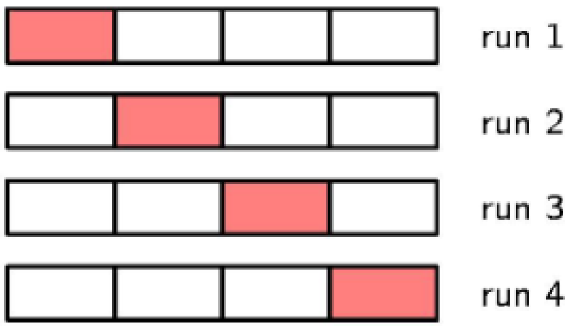
Validacijska zbirka za izbiro modela

- Izvedba na učni množici ni dober pokazatelj prediktivne zmogljivost
- Če je mnogo podatkov,
  - uporabimo del podatkov za učenje modelov ali določenega modela z izbranimi parametri
  - jih primerjamo z uporabo neodvisne validacijske zbirke
  - izberemo model z najboljšo prediktivno zmogljivost
- Če je zbirka podatkov majhna, lahko pride do ”over-fitting” pojava in je potrebno izbrati posebej testno množico

S-kratna križna validacija

- Dobava podatkov je omejena
- Vsi razpoložljivi podatki so razdeljeni v S skupin
- S-1 skupin uporabimo za učenje in evaluacijo na preostali skupini
- Postopek ponovimo za vseh S izbir skupin
- Rezultat je povprečna vrednost vseh S ponovitev

S=4



If S=N this is the leave-one-out method

Bayesov informacijski kriterij BIC

- Kriterij za izbiro modela
- *Akaike informacijski kriterij* (AIC)* izbere model, pri katerem je
- $\ln p(D|w_{ML}) - M$
  največja, pri tem je M število nastavljivih parametrov
- BIC** je različica te kvantitete

* The **Akaike information criterion** is a measure of the relative goodness of fit of a statistical model
In statistics, the **Bayesian information criterion (**BIC**) or **Schwarz criterion** (also **SBC**, **SBIC**) is a criterion for model selection among a finite set of models

Problem dimenzionalnosti

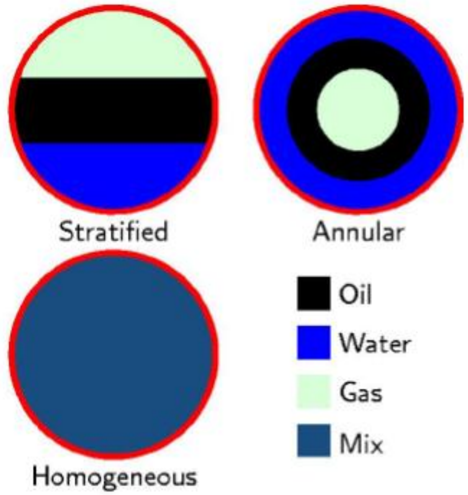
Rokovanje s prostori z več spremenljivkami in strojnim učenjem

Primer klasifikacije

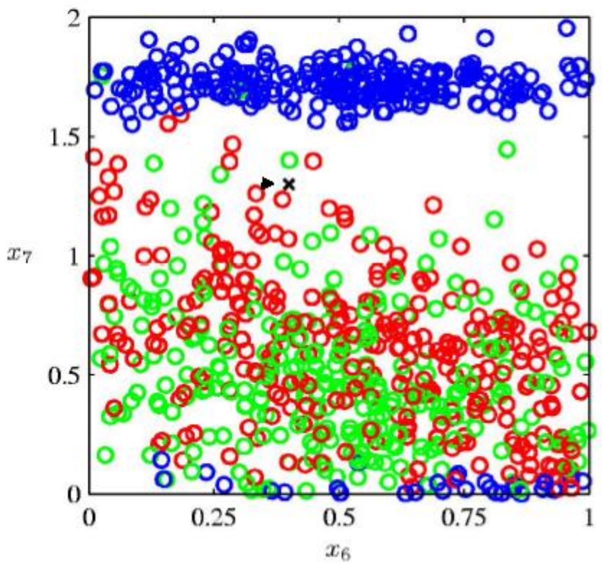
- Trije razredi
- 12 spremenljivk:

  2 sta prikazani
- 100 točk
- Naučiti se

  klasificirati podatke

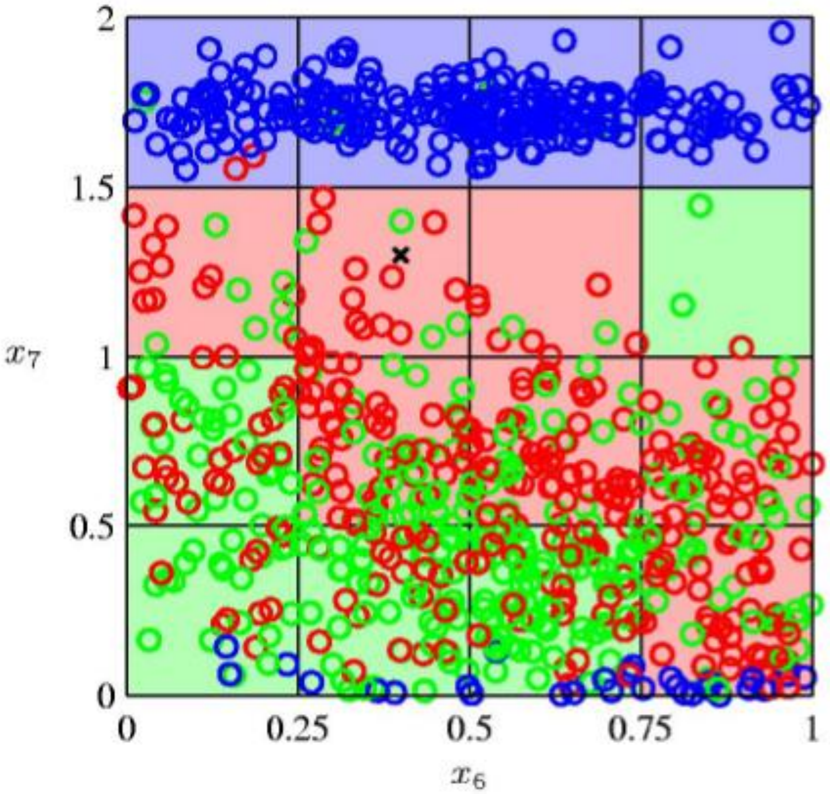


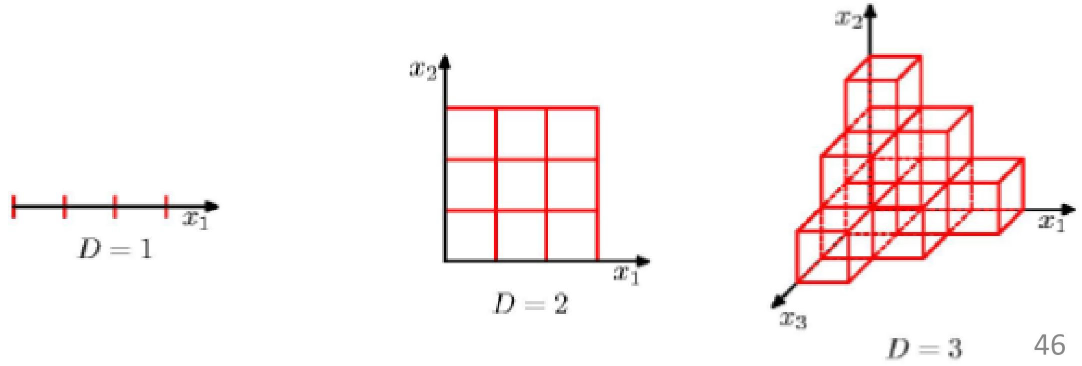
V kateri razred spada x?



Celična klasifikacija

- Preprosti pristop glasovanja na osnovi pripadnosti celici bo neuspešen zaradi eksponentnega naraščanja števila celic z naraščanjem dimenzije
- Malo točk v celicah





Volumen večdimenzionalne krogle

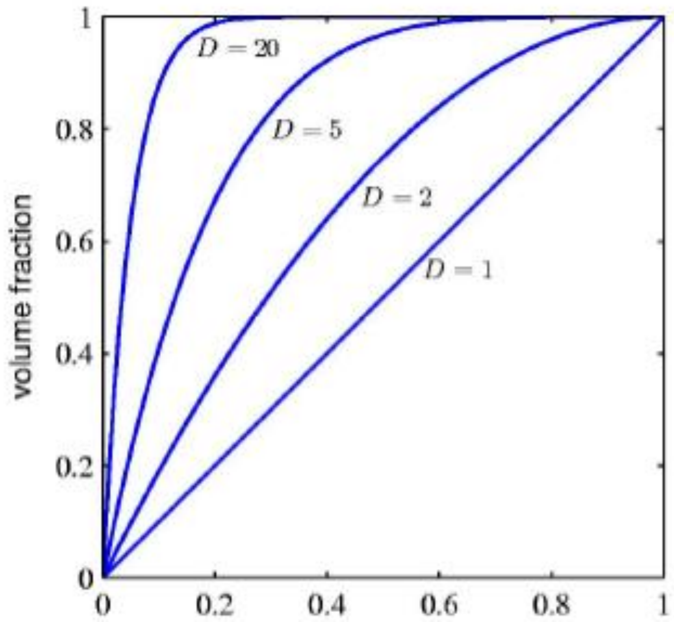
- Krogla ima polmer *r=1 v* *D*-dimenzijah
- Kolikšen del volumna leži

med polmeroma $r = 1- \sum$ in $r=1$?

- $V_D(r)=K_D r_D$
- Ta delež je določen z

  $1 - (1- \epsilon)^D$
- S povečevanjem *D* vse večji

delež volumna leži v bližini zunanje lupine



Fraction of volume of sphere lying in range r =1- ε to r = 1 for various dimensions D

Gauss v več-dimenzionalnem prostoru

- Prostor x-y preslikamo v prostor *r* z uporabo polarnih koordinat
- Večina porazdelitve je zgoščena na ozkem področju okrog določenega radija