

Sylabus

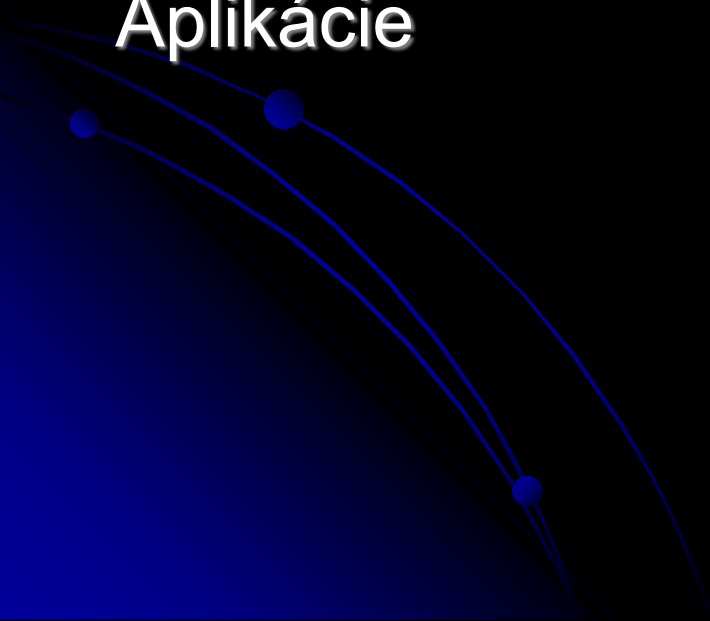
Príznaky, extrakcia (PV)

Príznaky - predspracovanie

Klasifikátory

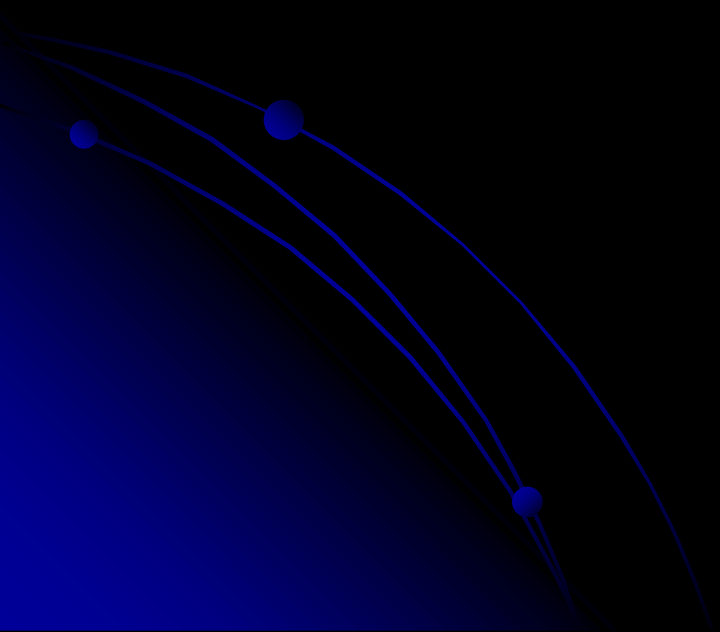
Hodnotenie klasifikátorov

Aplikácie



60 bodov skúška
40 bodov cvičenie

Z každého aspoň 50%



Clustering of microarray data.

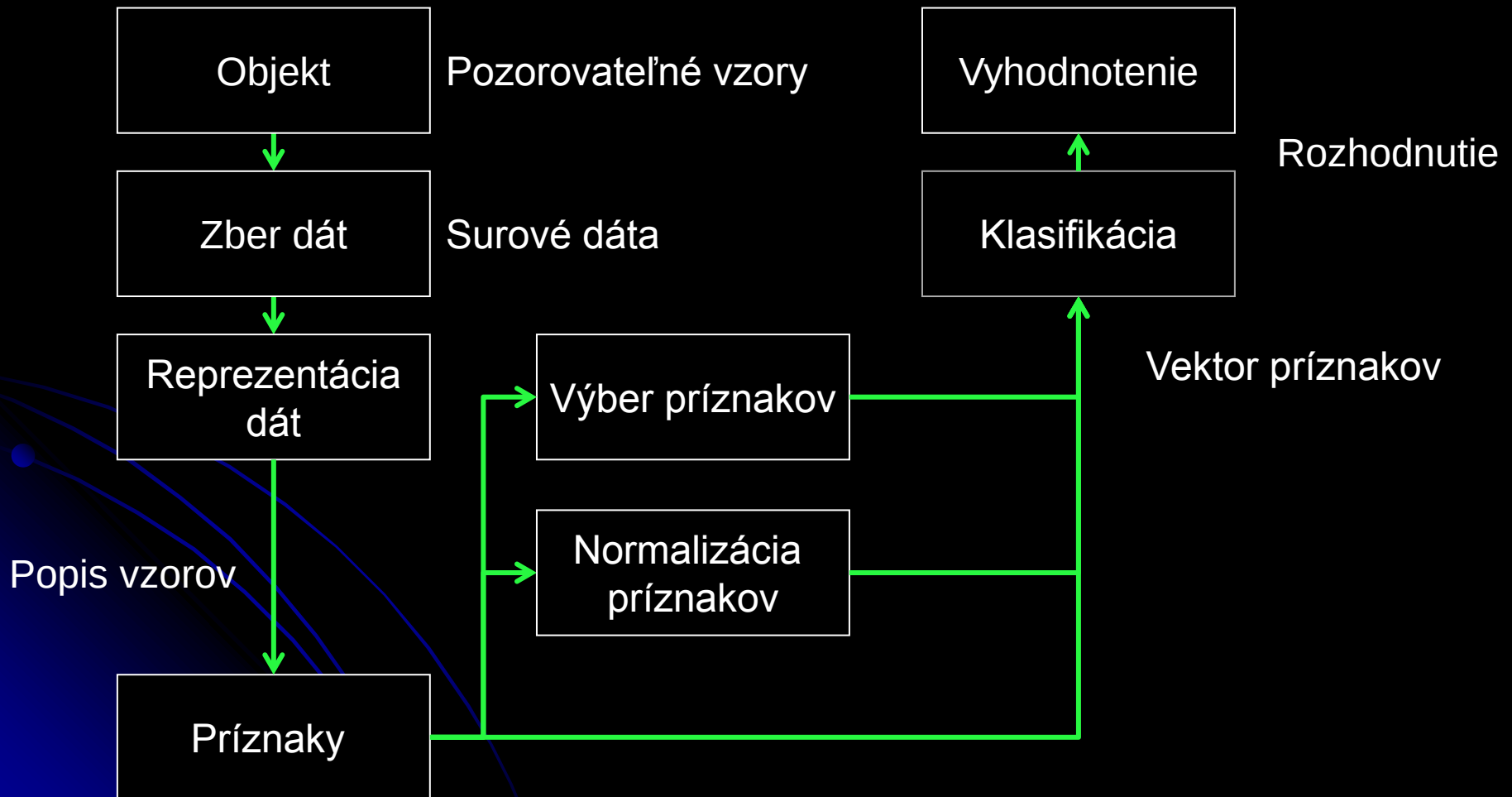
Pattern recognition

“The assignment of a physical object or event to one of several prespecified categories”

R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*

Problem Domain	Application	Input Pattern	Pattern Classes
Document image analysis	Optical character recognition	Document image	Characters, words
Document classification	Internet search	Text document	Semantic categories
Document classification	Junk mail filtering	Email	Junk/non-junk
Multimedia database retrieval	Internet search	Video clip	Video genres
Speech recognition	Telephone directory assistance	Speech waveform	Spoken words
Natural language processing	Information extraction	Sentences	Parts of speech
Biometric recognition	Personal identification	Face, iris, fingerprint	Authorized users for access control
Medical	Computer aided diagnosis	Microscopic image	Cancerous/healthy cell
Military	Automatic target recognition	Optical or infrared image	Target type
Industrial automation	Printed circuit board inspection	Intensity or range image	Defective/non-defective product
Industrial automation	Fruit sorting	Images taken on a conveyor belt	Grade of quality
Remote sensing	Forecasting crop yield	Multispectral image	Land use categories
Bioinformatics	Sequence analysis	DNA sequence	Known types of genes
Data mining	Searching for meaningful patterns	Points in multidimensional space	Compact and well-separated clusters

Rozpoznávanie



Príznamy

Obchodný dom:

Nohavice, košele – *odevy*

Rajčiny, šalát – *zelenina*

Pomaranče, jablká – *ovocie*

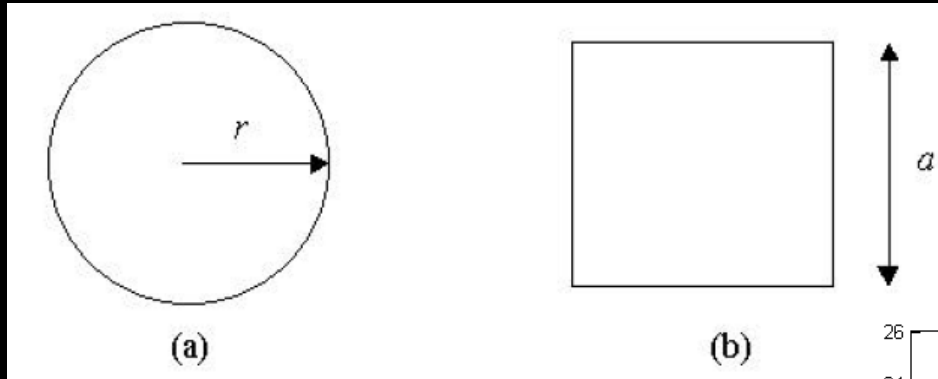
alternatívne kritériá – rôzna klasifikácia

“zelené veci”

jablká, šalát aj košele v zelenom oddelení

Kritérium výberu príznakov ovplyvňuje klasifikáciu

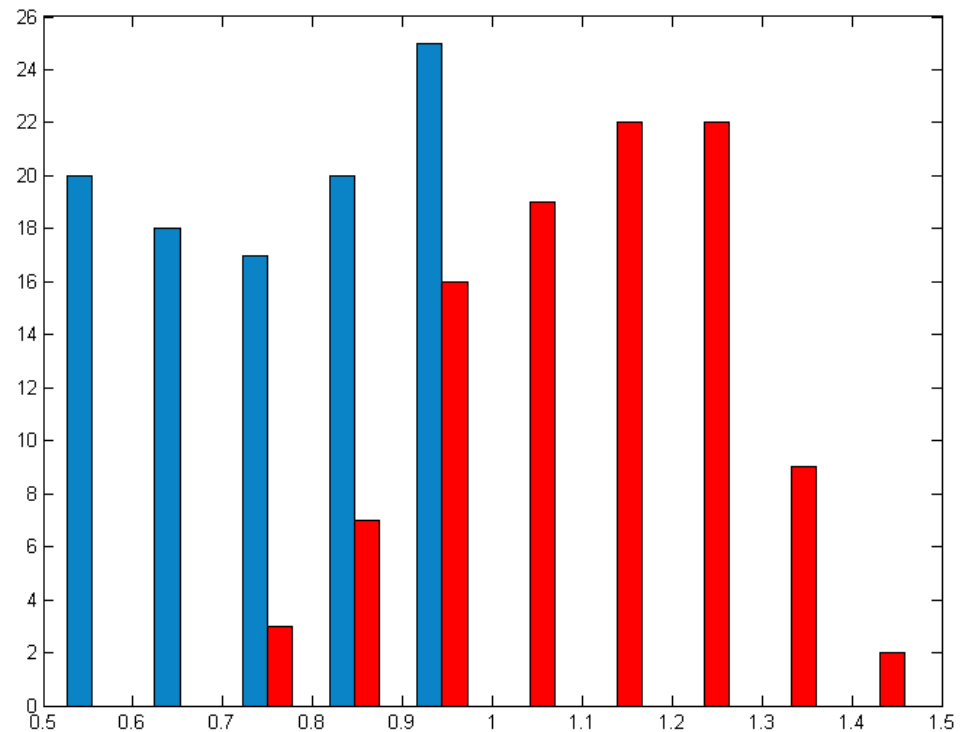
Príklad



Cukríky v krabiciach

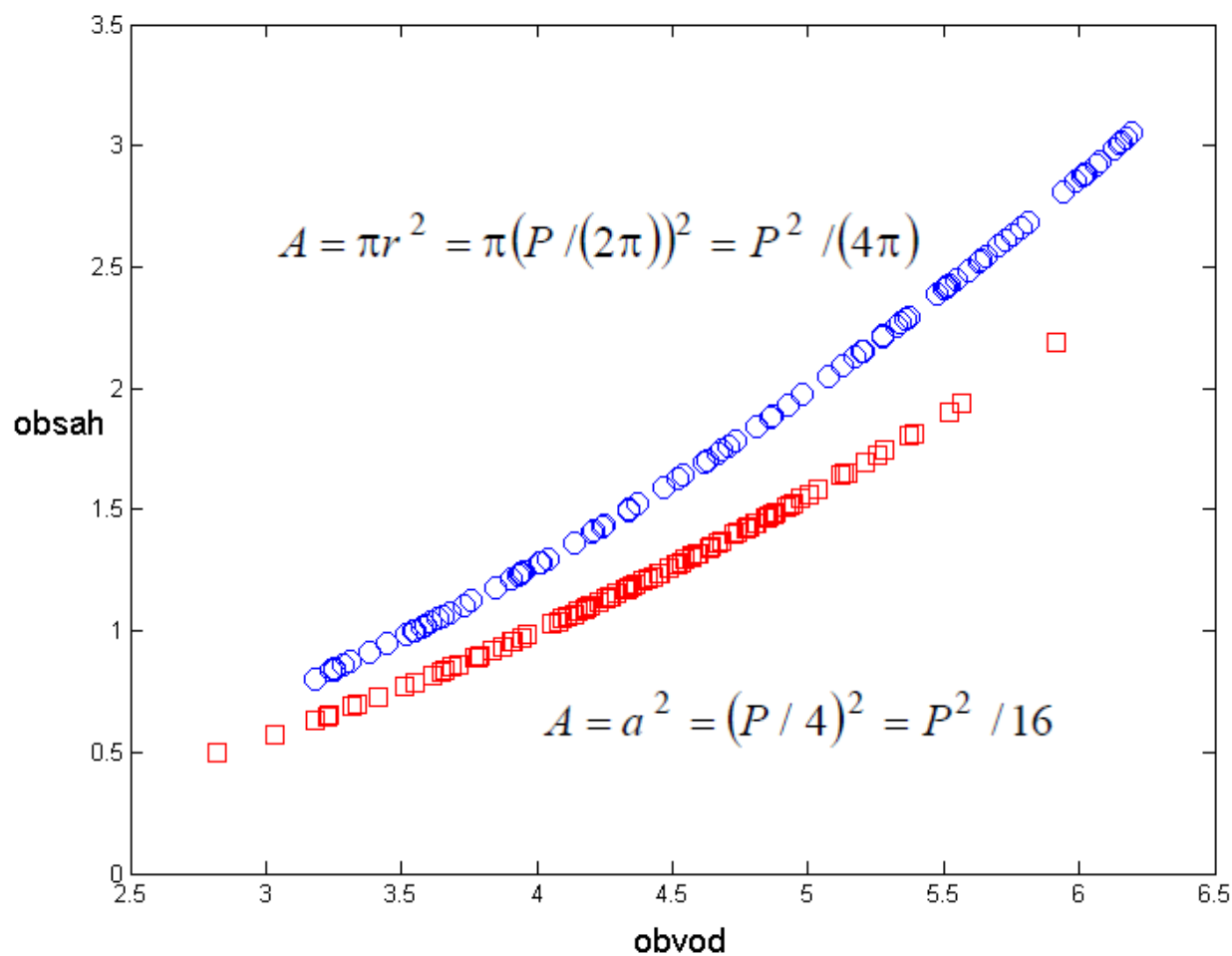
$$0.7 < a \leq 1.5$$

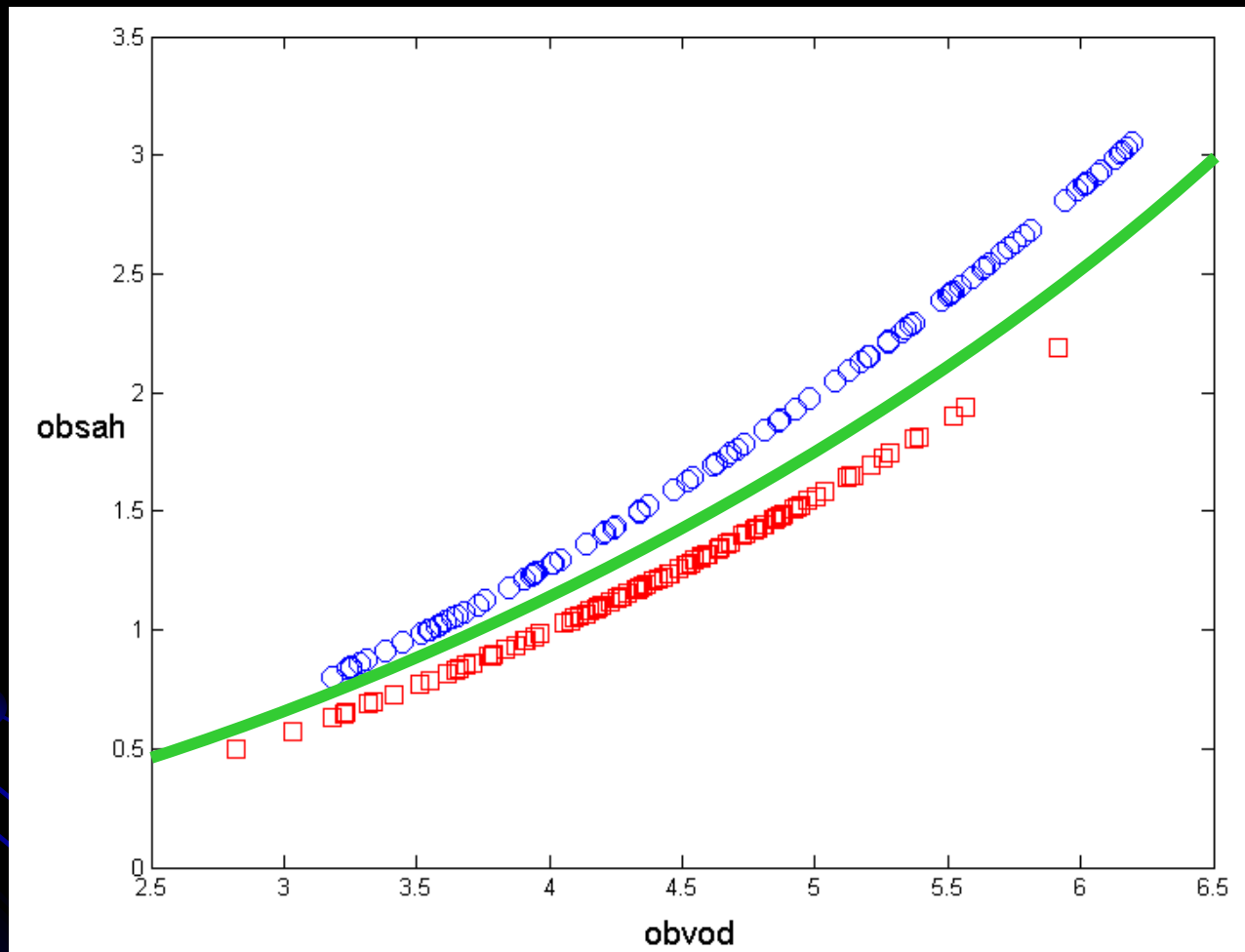
$$0.5 < r \leq 1$$



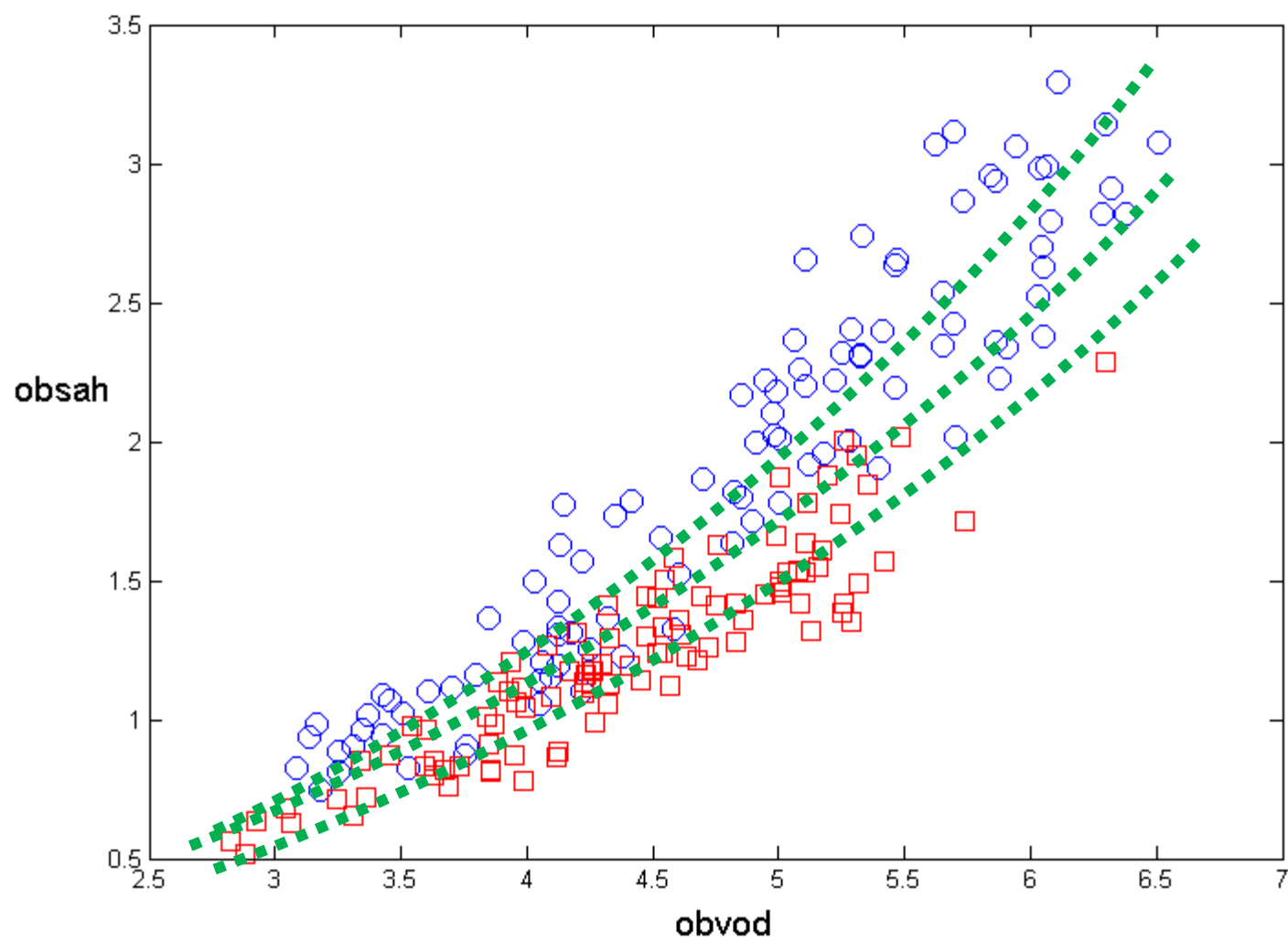
	kruh	štvorec
obvod	$P(r) = 2\pi r$	$P(r) = 4a$
obsah	$A(r) = \pi r^2$	$A(r) = a^2$

(*obvod x obsah*) priestor príznačov

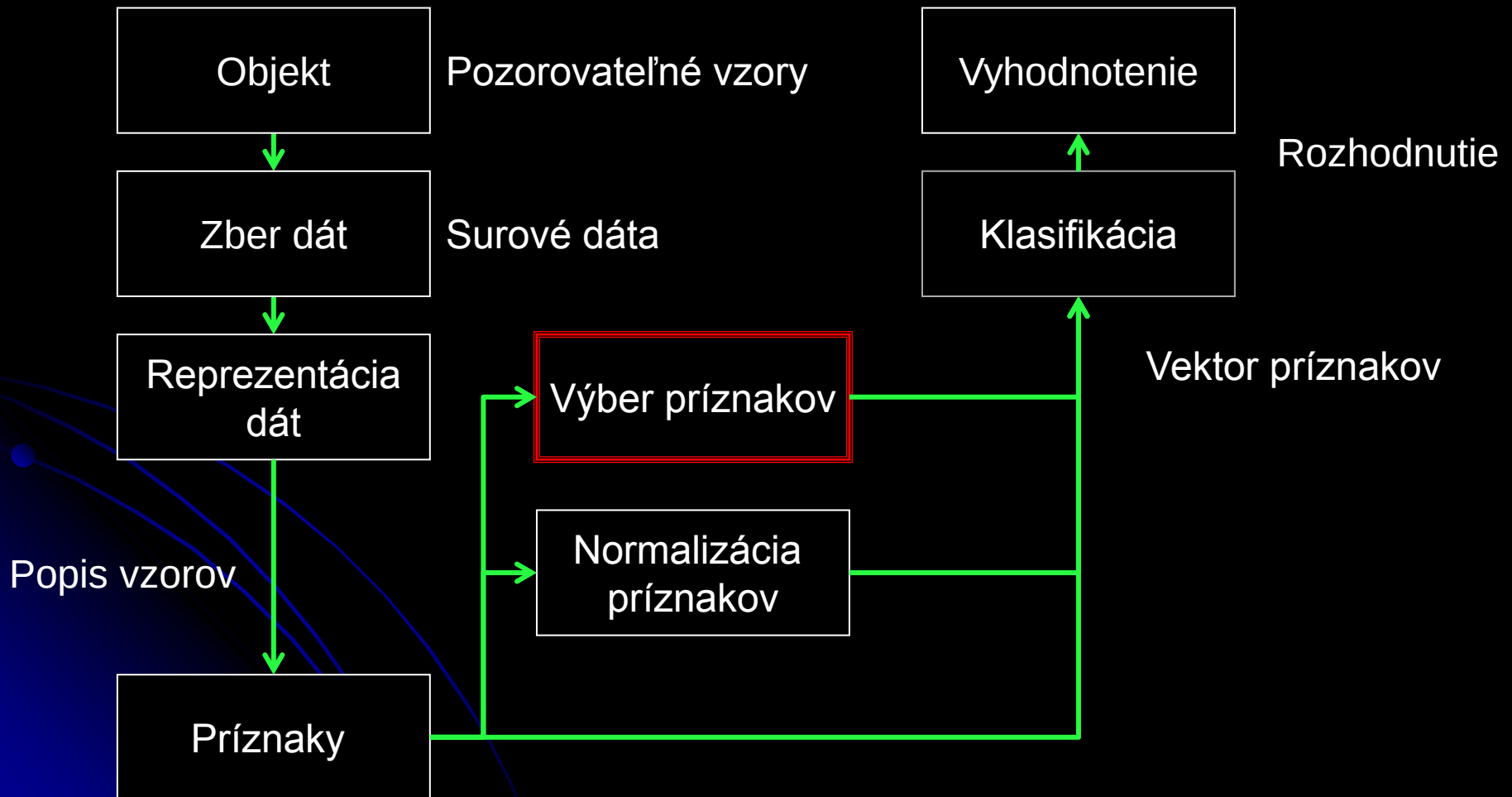




$$A = P^2 / k, \text{ where } 4\pi < k < 16$$



Rozpoznávanie

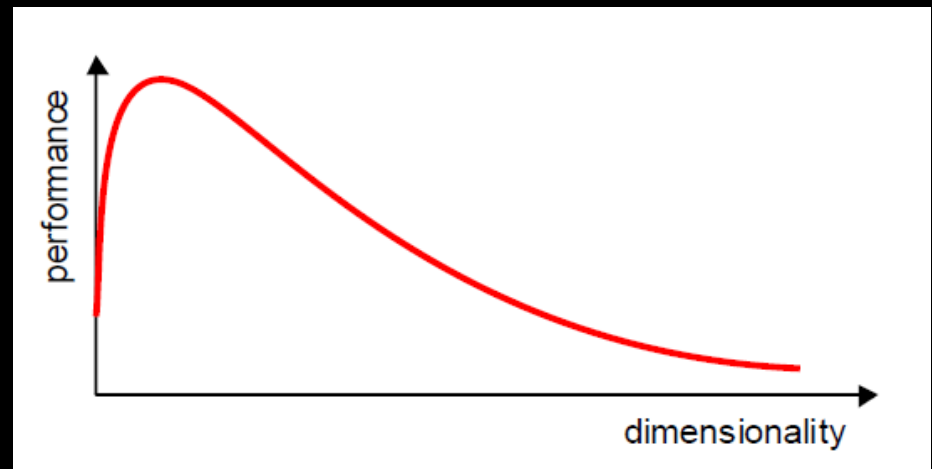


Príznaky

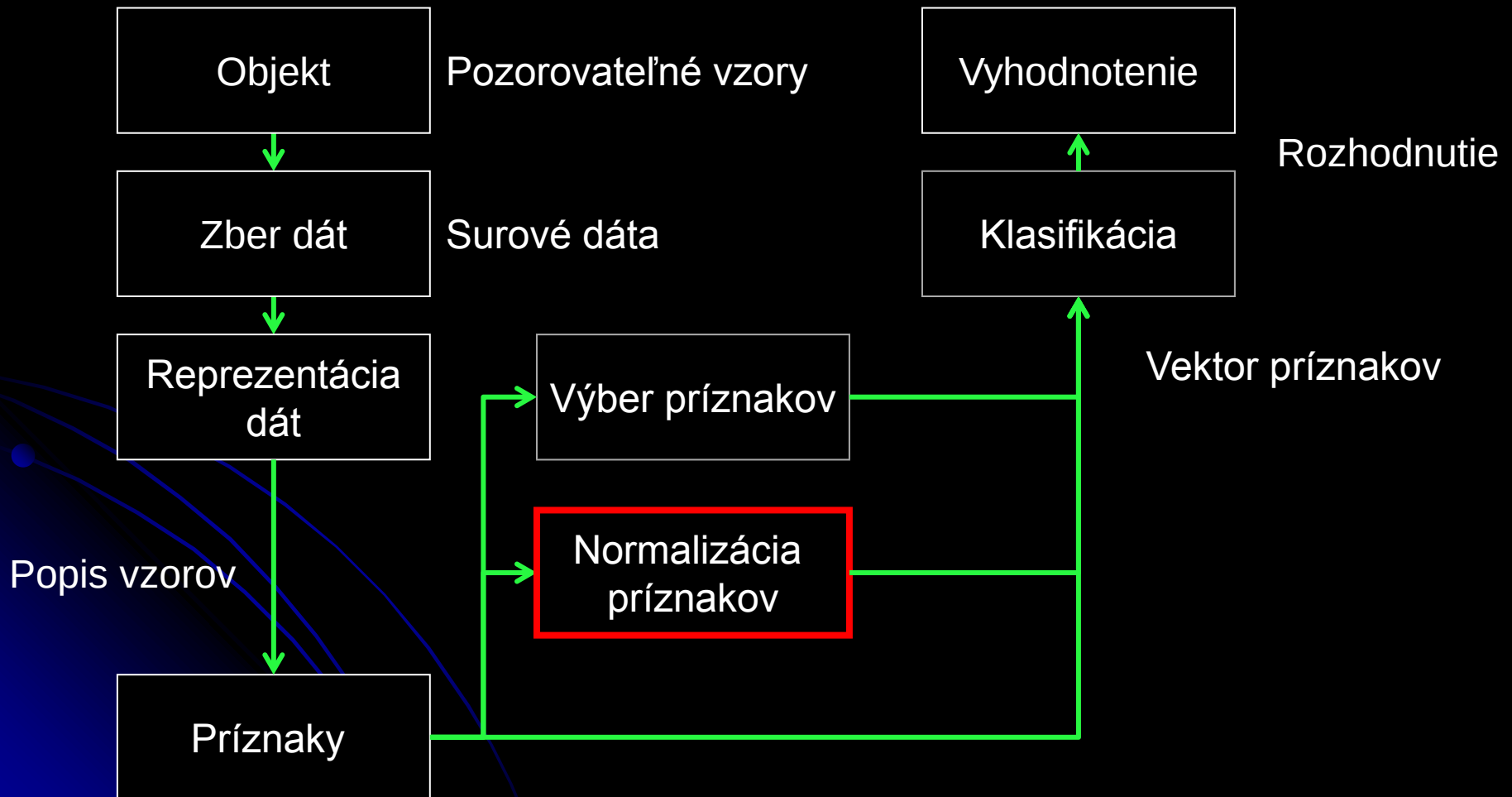
- viac príznakov => viac informácie, vyššia presnosť
- viac príznakov => zložitejšia extrakcia
- viac príznakov => zložitejší tréning klasifikátora

The curse of dimensionality – preklatie dimenzionality

Riešenie:
Optimálny počet príznakov

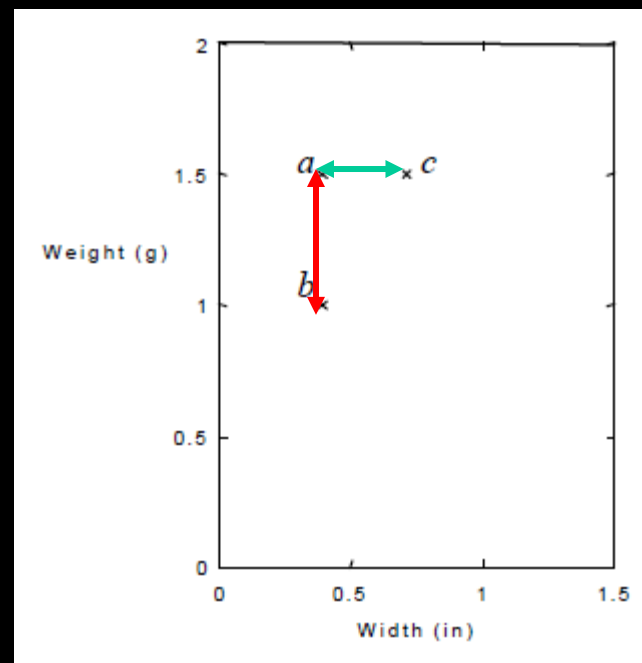
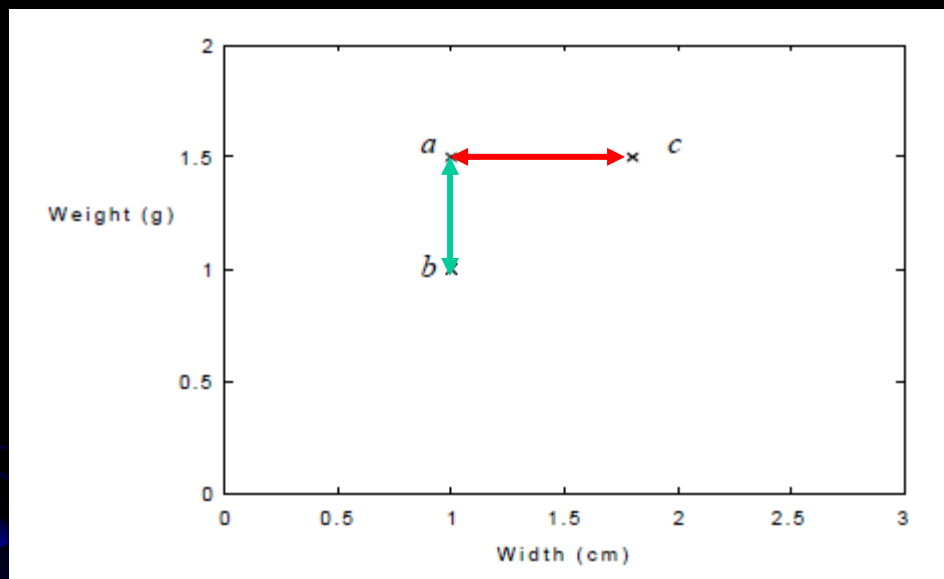


Rozpoznávanie



Normalizácia príznačov

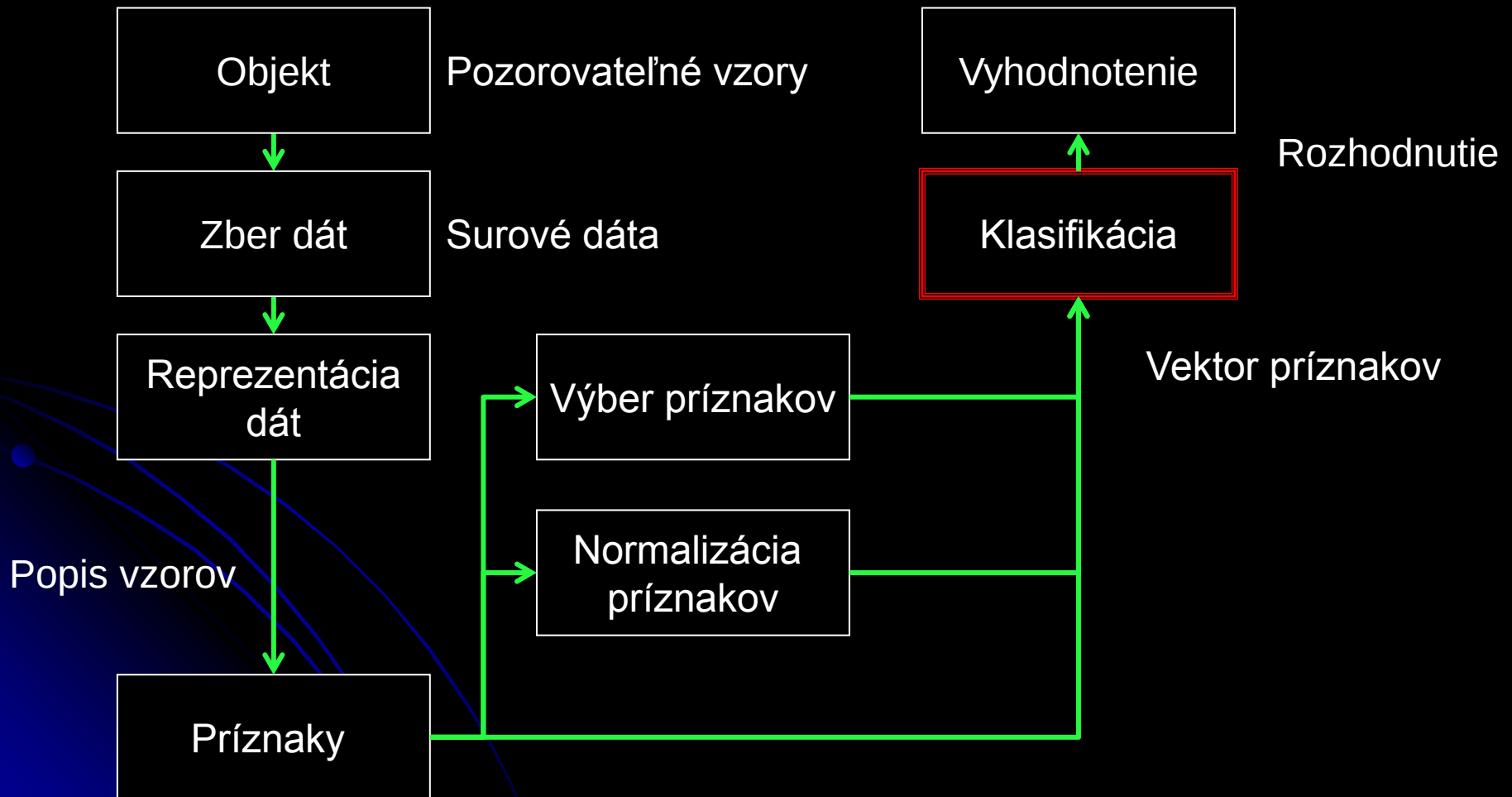
Príznaky – rozmer (cm, m, cl, ml, ...)

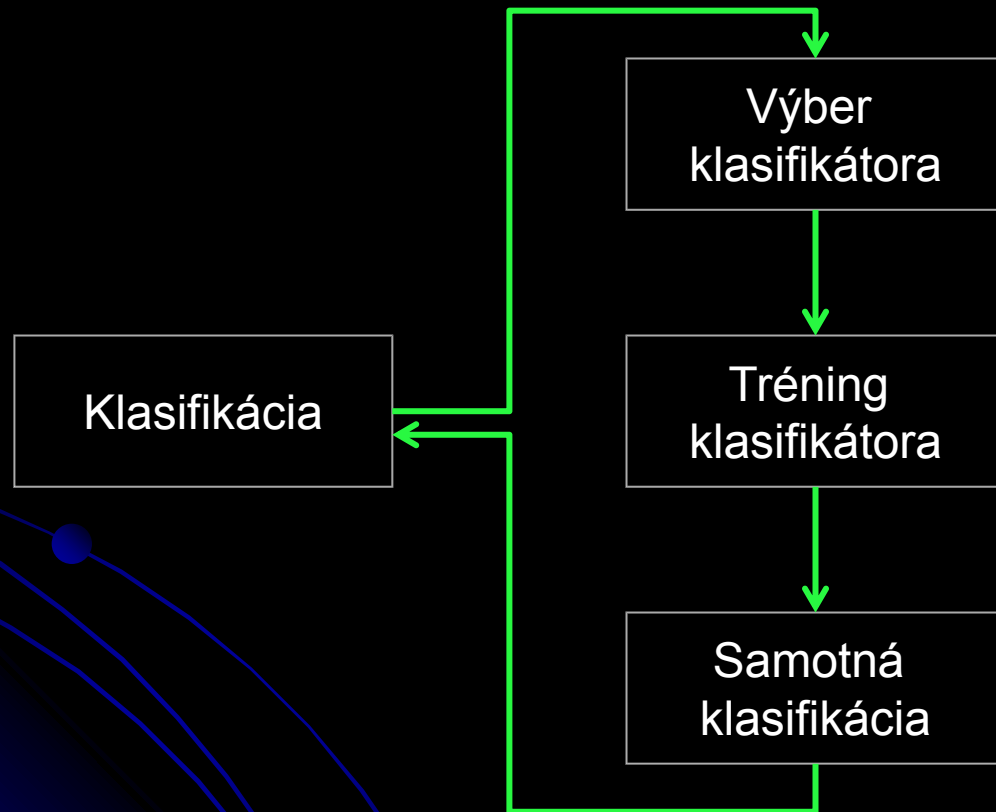


Výber jednotiek ovplyvňuje vzdialenosť v priestore príznačov

Normalizáciou získame bezrozmerné príznaky

Rozpoznávanie







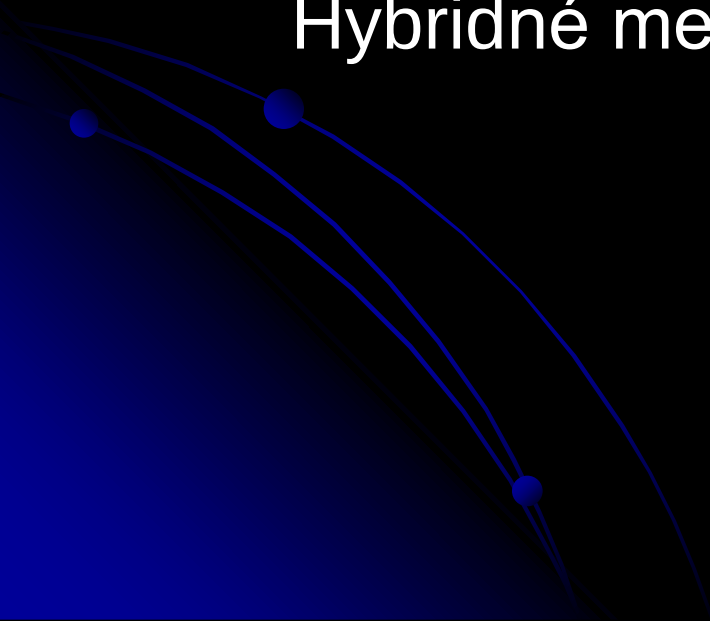
Typy klasifikátorov

Šablóny (PV)

Príznakové metódy

Syntaktické (štrukturálne) metódy

Hybridné metódy



Príznakové metódy

Štatistika

Bayesovská teória rozhodovania

Vzdialenosti medzi
príznakmi

techniky najbližších susedov

diskriminačná analýza

mechanizmy podporných vektorov

metóda K priemerov

hierarchické zhľukovanie

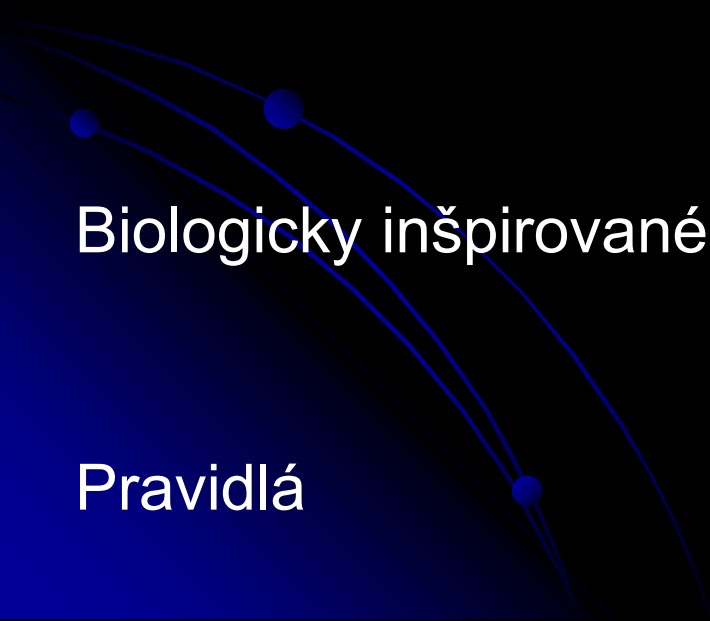
Biologicky inšpirované

neurónové siete

samoorganizujúce mapy

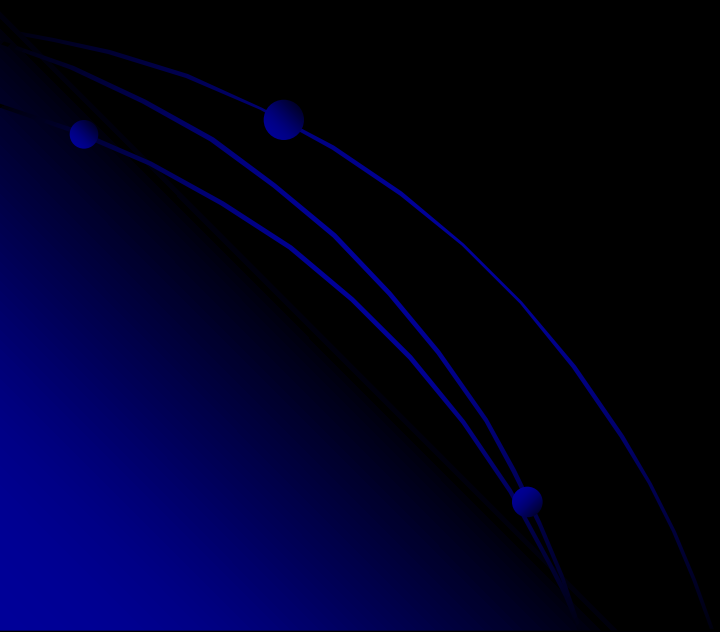
Pravidlá

rozhodovacie stromy



Klasifikácia – komplexný proces

výber príznakov, mier, kritérií, parametrov.....



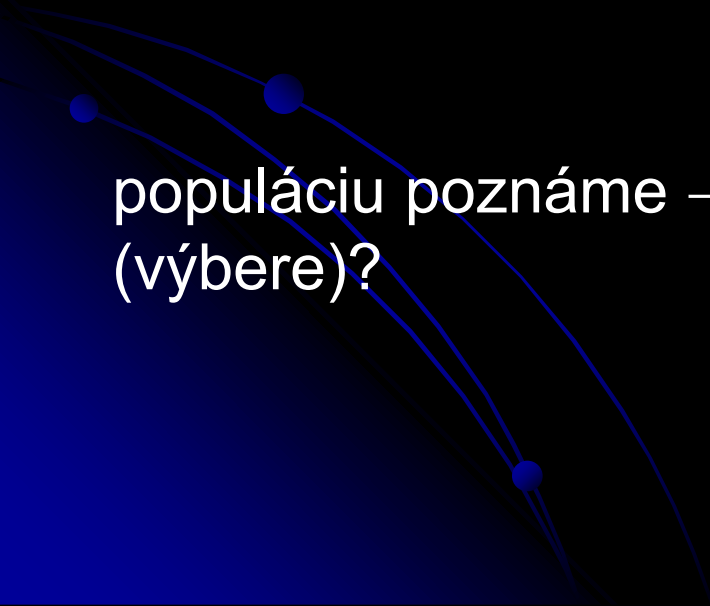


Pravdepodobnosť

- Abstraktný matematický model neurčitosti
- Modeluje deje, v ktorých hrá rolu náhodnosť

Dedukcia: usudzovanie zo všeobecného na konkrétne

populáciu poznáme → čo môžeme povedať o vzorke (výbere)?



Štatistika

- Zber, organizácia a analýza dát
- Zobecňuje z obmedzených / konečných vzoriek
- Odhad parametrov, testovanie hypotéz, atď

Indukcia: usudzovanie z konkrétneho na všeobecné

- vzorku poznáme → čo môžeme povedať o populácii?

drvivá väčšina poznania stojí na indukcii
(experimentálne vedy)

Ak sme už kúpili niekoľko rovnakých výrobkov a niektoré z nich sa pokazili, čo vieme povedať o tomto type výrobku vo všeobecnosti?

- Ak poznáme kazivosť daného typu výrobku, čo vieme o konkrétnom výrobku, ktorý si kupujeme?

História pravdepodobnosti

Úspech v hazardných hrách

Hra: Hod férovou kockou

Ak padne 2, 3 alebo 4, vyhráte 1 Euro

Ak padne 5, vyhráte 2 Eurá

Ak padne 1 alebo 6, prehráte 3 Eurá

Kto by hral túto hru?



Definícia pravdepodobnosti

Klasická (Laplaceova)

relatívna početnosť výsledku A v N pokusoch

$$P(A) = \frac{N_A}{N}$$

Limitná

relatívna početnosť v nekonečne veľkom počte pokusov

$$P(A) = \lim_{N \rightarrow \infty} \frac{N_A}{N}$$

Axiomatická (Kolmogorovova)

Pole javov $\mathcal{A}$ je σ -algebra na Ω ,

t.j. systém podmnožín priestoru Ω obsahujúci Ω a uzavretý voči doplnku a zjednoteniu

$$\begin{aligned}\Omega &\in \mathcal{A}, \\ A \in \mathcal{A} &\Rightarrow A^C \in \mathcal{A}, \\ (\forall n \in \mathbb{N}: A_n \in \mathcal{A}) &\Rightarrow \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}.\end{aligned}$$

Pravdepodobnosť je funkcia $P: \mathcal{A} \rightarrow \langle 0,1 \rangle$, ktorá spĺňa

$$\begin{aligned}P(\Omega) &= 1, \\ P\left(\bigcup_{n \in \mathbb{N}} A_n\right) &= \sum_{n \in \mathbb{N}} P(A_n)\end{aligned}$$

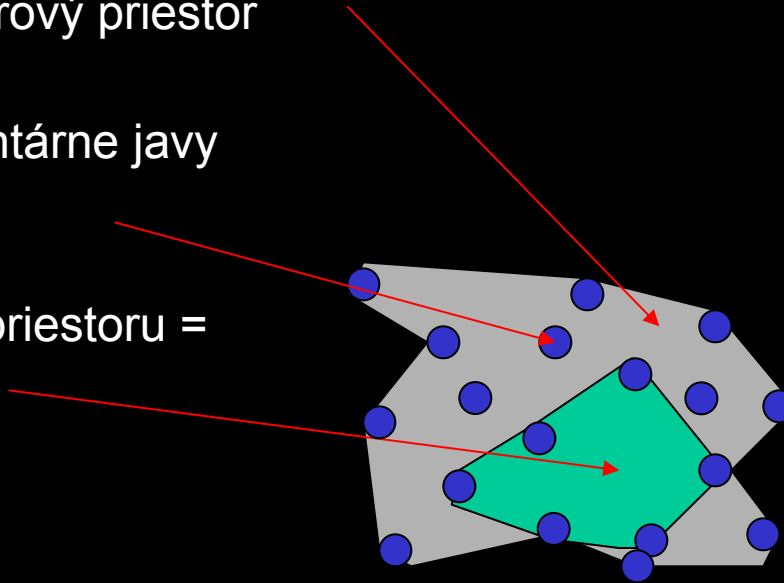
ak javy A_n sú po dvoch disjunktné

Pravdepodobnostný priestor $(\Omega, \mathcal{A}, P)$

Množina všetkých možných výstupov =
výberový priestor

Jednotlivé prvky = elementárne javy

Podmnožina výberového priestoru =
(náhodný) jav



Jav:

Diskrétny (napr. hod kockou alebo mincou, počet detí v rodine, ...)

Spojité (napr. rozmer nejakého výrobku, výška hráčov, ...)

Kvantitatívny (vyjadruje množstvo – 17, 42, 10001, ...)

Kvalitatívny (vyjadruje kvalitu – červený, Škoda, Slovensko, ...)

Príklady

Množina všetkých možných výstupov = výberový priestor

Jednotlivé prvky = elementárne javy

Podmnožina výberového priestoru = (náhodný) jav

3 hody mincou

Výberový priestor Ω

$\{HHH, HHO, HOH, HOO, OHH, OHO, OOH, OOO\}$.

Javy:

$A = \{\text{aspoň 2x Hlava}\}$
 $= \{HHH, HHO, HOH, OHH\}$

$B = \{\text{práve 2x Orol}\}$
 $= \{HOO, OHO, OOH\}$

1 hod kockou

Výberový priestor Ω

$\{1, 2, 3, 4, 5, 6\}$.

Javy:

$A = \{\text{párne čísla}\}$
 $= \{2, 4, 6\}$

$B = \{\text{čísla menšie ako 5}\}$
 $= \{1, 2, 3, 4\}$

Zjednotenie A a B , $A \cup B$, nastáva práve vtedy, keď nastane aspoň jeden z javov A a B

Opačný jav k javu A , A^c , sa nazýva aj jav doplnkový alebo komplementárny ($\Omega \setminus A$).

Prienik A a B , $A \cap B$, nastáva práve vtedy, keď nastanú oba javy A a B

Keď javy A a B nemajú ani jeden spoločný elementárny jav, sú nezlučiteľné (nemôžu nastať oba naraz)

• **Rozkladom výberového priestoru** nazývame súbor vzájomne disjunktných neprázdnych javov, ktorých zjednotením je celý výberový priestor.

$$\{A_i\}_{i=1}^M$$

$$\forall i, j: A_i \cap A_j = \emptyset$$

$$\bigcup_{i=1}^M A_i = \Omega$$

Počet Hláv v 10 hodoch mincou

$$A = \{0, 2, 4, 6, 8, 10\}$$

$$B = \{1, 3, 5, 7, 9\}$$

$$C = \{0, 1, 2, 3, 4, 5\}$$

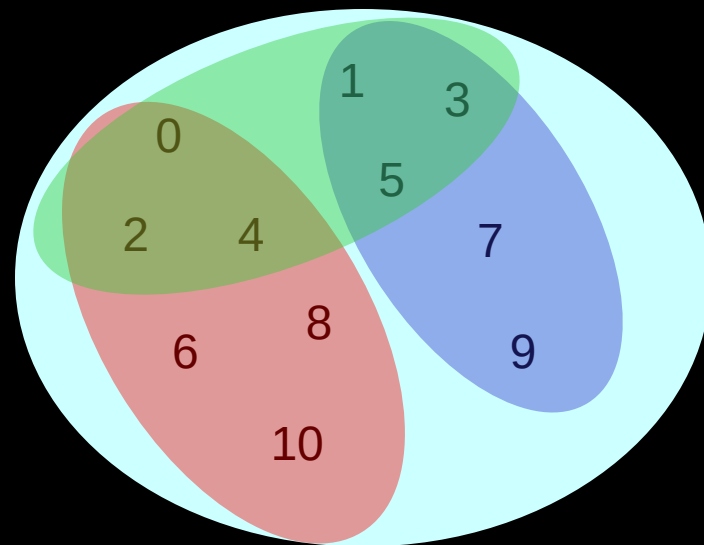
$$A \cup B = \{0, 1, \dots, 10\} = \Omega$$

$$A \cap B = \emptyset$$

=> rozklad Ω

$$C^c = \{6, 7, 8, 9, 10\}$$

$$A \cap C = \{0, 2, 4\}.$$



vlastnosti

$$P(\emptyset) = 0$$

$$P(A^C) = 1 - P(A)$$

$$A \subseteq B \Rightarrow P(A) \leq P(B)$$

$$P(B \setminus A) = P(B) - P(A \cap B)$$

$$A \subset B \Rightarrow P(B \setminus A) = P(B) - P(A)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$$

$$P(A \cup B) = 1 - P((A \cup B)^C) = 1 - P(A^C \cap B^C)$$

$$P(A \cap B) = 1 - P((A \cap B)^C) = 1 - P(A^C \cup B^C)$$

Javy A a B nazveme nezávislé

$$P(A \cap B) = P(A) \cdot P(B)$$

Aká je pravdepodobnosť, že pri hode mincou padne hlava dvakrát za sebou?

Pri férovej minci je pravdepodobnosť,
že padne hlava 0,5.

Výsledok druhého hodu (B) je nezávislý
od výsledku prvého hodu (A).

Preto výsledná pravdepodobnosť
 $P(A \cap B) = P(A) \cdot P(B) = 0,5 \cdot 0,5 = 0,25$.

Podmienená pravdepodobnosť

pravdepodobnosť, že nastane jav A za podmienky, že nastal jav B

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$A = \{0, 2, 4, 6, 8, 10\}$$

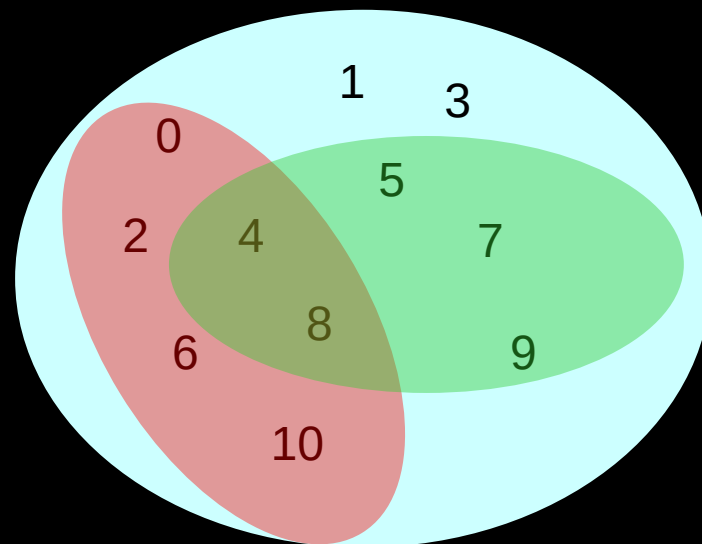
$$B = \{4, 5, 7, 8, 9\}$$

$$A \cap B = \{4, 8\}$$

$$P(A \cap B) = 2/11$$

$$P(B) = 5/11$$

$$P(A|B) = 2/5$$



*Z krabice, v ktorej sa nachádza 10 červených a 5 modrých ponožiek vytiahneme naslepo postupne dve ponožky.
Aká je pravdepodobnosť, že sme ako druhú vytiahli modrú ponožku?*

nejedná sa o nezávislé javy

Označme si

A = jav, keď bola vytiahnutá ako prvá červená ponožka

B = jav, keď bola vytiahnutá ako prvá modrá ponožka

C = vytiahnutie modrej ponožky v druhom ťahu

$$P(A) = 10/15$$

$$P(B) = 5/15$$

$$P(C) = P((C \cap A) \cup (C \cap B))$$

$$P(C) = P(C \cap A) + P(C \cap B)$$

$$P(C) = P(C \cap A) + P(C \cap B)$$

$$P(C|A) = P(C \cap A)/P(A)$$

$$P(C|B) = P(C \cap B)/P(B)$$

$$P(C) = P(C|A).P(A) + P(C|B).P(B) = \frac{5}{14} \cdot \frac{10}{15} + \frac{4}{14} \cdot \frac{5}{15} = \frac{7}{21} = \frac{1}{3}$$

$$\{A_i\}_{i=1}^M$$

rozklad priestoru $\Omega \Rightarrow$

$$\sum_{i=1}^M P(A_i) = 1$$

Potom

$$P(B) = \sum_{i=1}^M P(B|A_i)P(A_i)$$

Veta o úplnej pravdepodobnosti

Ľubovoľný jav $B \in \Omega$ je zjednotením javov $(B \cap A_1), \dots, (B \cap A_M)$

$$B = \bigcup_{i=1}^M (B \cap A_i)$$

Tieto javy sú disjunktné, preto $P(B) = \sum_{i=1}^M P(B \cap A_i)$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(B) = \sum_{i=1}^M P(B|A_i)P(A_i)$$

ktorý z javov $\{A_i\}$ viedol k javu B?

$$P(A_k|B)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(B) = \sum_{i=1}^M P(B|A_i)P(A_i)$$

$$\begin{aligned} P(A_k|B) &= \frac{P(B|A_k) \cdot P(A_k)}{P(B)} \\ &= \frac{P(B|A_k) \cdot P(A_k)}{\sum_{i=1}^M P(B|A_i)P(A_i)} \end{aligned}$$

Bayesov vzorec

Na konci výrobnéj linky je tester, ktorý s pravdepodobnosťou 0,999 vyradí chybný výrobok.

S pravdepodobnosťou 0,01 však vyradí aj kvalitný výrobok.

Kazivosť výrobkov je 0,1 (jeden z desiatich výrobkov je chybný).

Aká je pravdepodobnosť, že vyradený výrobok je skutočne chybný?

$$P(A_{ch}) = 0,1$$

$$P(A_{kv}) = 0,9$$

Vyradenie výrobku – jav B

$$P(B|A_{ch}) = 0,999$$

$$P(B|A_{kv}) = 0,01$$

$$\begin{aligned} P(A_{ch}|B) &= \frac{P(B|A_{ch}) \cdot P(A_{ch})}{P(B|A_{ch})P(A_{ch}) + P(B|A_{kv})P(A_{kv})} \\ &= \frac{0,999 \cdot 0,1}{0,999 \cdot 0,1 + 0,01 \cdot 0,9} \end{aligned}$$