# OSNOVE UMETNE INTELIGENCE

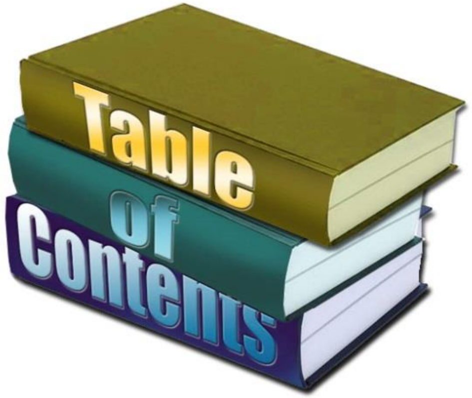
## 2021/22

*ocenjevanje učenja,*
*diskretizacija, manjkajoči atributi*
*naivni Bayesov klasifikator*

# Pridobljeno znanje s prejšnjih predavanj

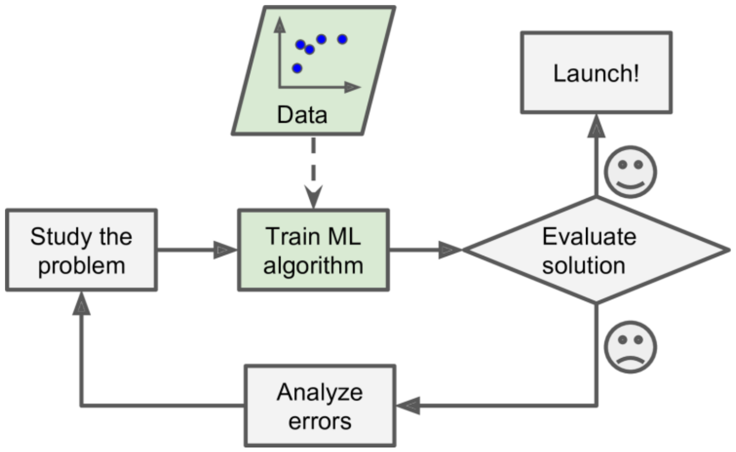
- **strojno učenje**
  - prostor hipotez odločitvenih dreves
  - problem pretiranega prileganja učnim podatkom
    - učenje šuma
    - prevelika drevesa, slaba razumljivost
    - slaba splošnost
  - **strategije rezanja**: vnaprej, nazaj
  - ocenjevanje verjetnosti z bolj stabilnimi ocenami
  - rezanje nazaj:
    - REP: **uporaba rezalne množice**, odrežemo poddrevesa, ki povečujejo število napak glede na število napak v njihovem korenu
    - MEP: **uporaba m-ocene verjetnosti**, ki kombinira apriorno in empirično verjetnost, ocenimo statično in dinamično napako

# Pregled




- strojno učenje
  - uvod v strojno učenje
  - učenje odločitvenih dreves
  - učenje dreves iz šumnih podatkov (rezanje dreves)
  - ocenjevanje učenja
  - diskretizacija atributov, obravnava manjkajočih vrednosti
  - naivni Bayesov klasifikator

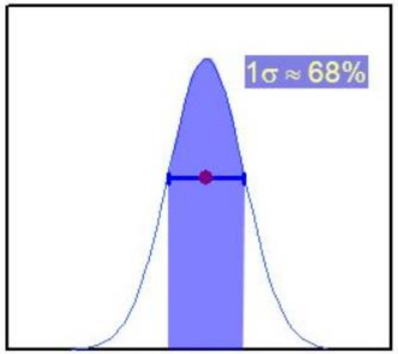
# Ocenjevanje učenja

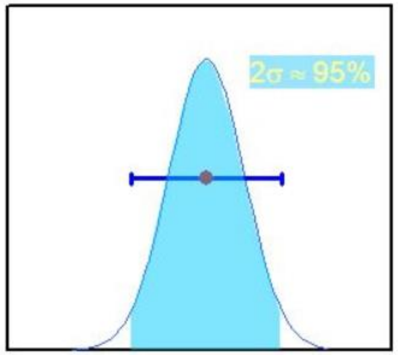



- kriteriji za ocenjevanje hipotez:
  - **točnost** (angl. *accuracy*) (**konsistentnost**, **splošnost**)
  - **razumljivost** (angl. *comprehensibility*) – subjektivni kriterij ali tudi **kompleksnost** (angl. *complexity*)

- ocenjevanje točnosti:
  - na **učnih** podatkih (angl. *training set, learning set)*
  - na **testnih** podatkih (angl. *testing set, test set*)
    - izločimo del učnih podatkov, s katerimi simuliramo ne-videne podatke
    - želimo si, da je testna množica reprezentativna za nove podatke
    - uporabimo lahko **intervale zaupanja** v oceno uspešnosti na testni množici, ki upoštevajo število testnih primerov
  - na **novih** (ne-videnih) podatkih (angl. *new data, unseen data*)
    - na njih bo naučeni sistem dejansko deloval

# Ocenjevanje učenja





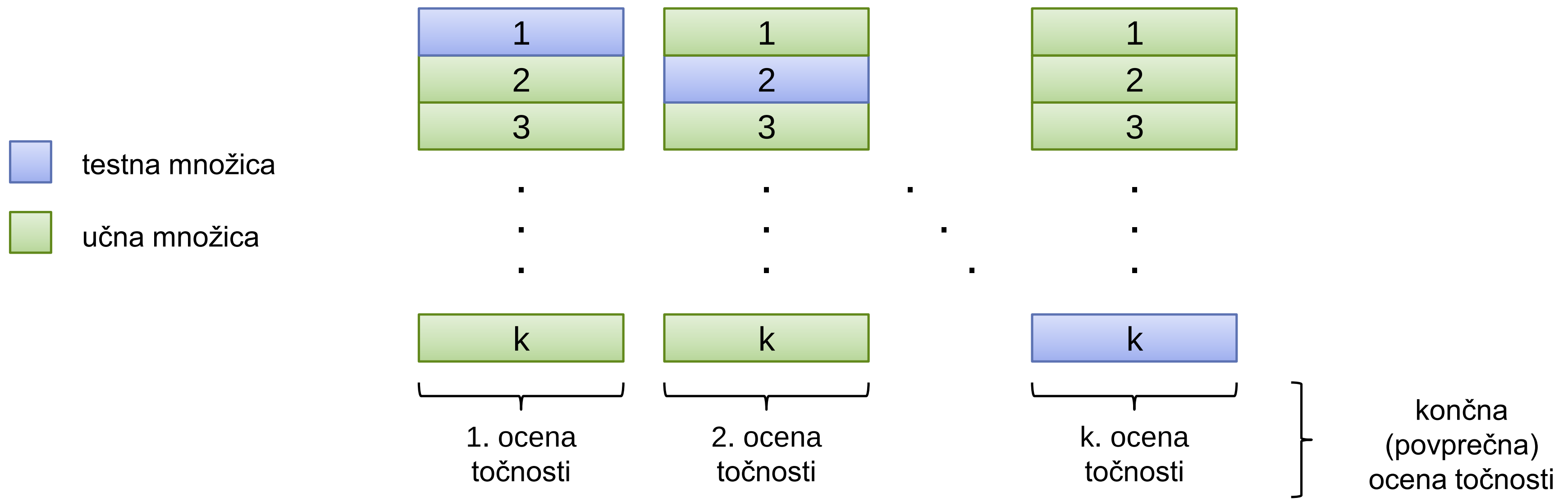

- nasprotujoča si cilja:
  - potrebujemo čim več podatkov za **uspešno učenje**
  - potrebujemo čim več podatkov za **zanesljivo ocenjevanje točnosti** (večje število testnih primerov nam daje ožji interval zaupanja v oceno točnosti)

- rešitev:
  - kadar je učnih podatkov dovolj, lahko izločimo **testno množico** (angl. *holdout test set*)
  - alternativa: **večkratne delitve** na učno in testno množico

- različni načini **vzorčenja testnih primerov**:
  - naključno, nenaključno (npr. prečno preverjanje)
  - poljubno ali stratificirano (zagotovimo enako porazdelitev razredov kot v učni množici)

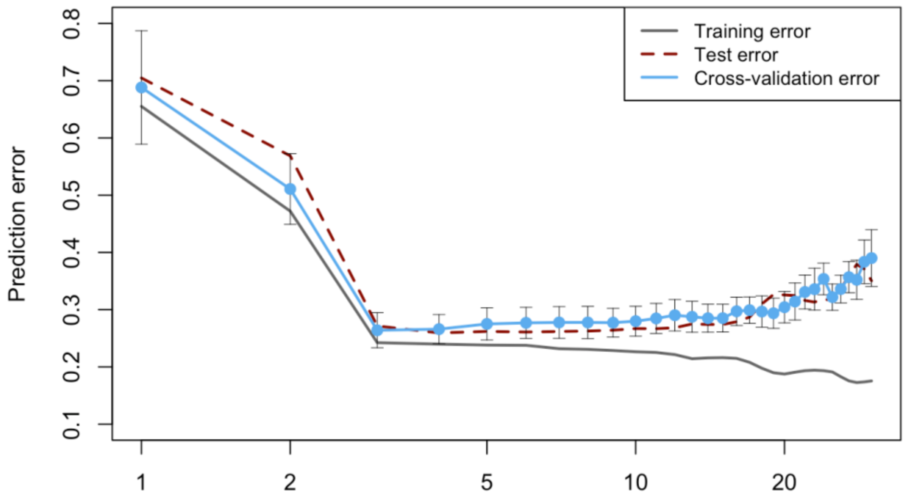
# Prečno preveranje

- poseben primer **večkratnega učenja** in testiranja
- **k-kratno prečno preverjanje** (angl. *k-fold cross-validation*):
  - celo učno množico razbij na *k* disjunktnih podmnožic
  - za vsako od *k* podmnožic:
    - uporabi **množico** kot **testno množico**
    - uporabi **preostalih *k-1*** množic kot **učno množico**
  - povpreči dobljenih *k* ocen točnosti v končno oceno

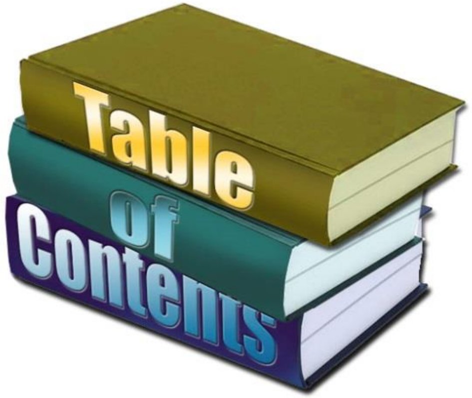


# Prečno preveranje

- v praksi **najpogosteje: k=10** (10-kratno prečno preverjanje)
- vplive izbranega razbitja podatkov na podmnožice lahko zmanjšamo tako, da tudi prečno preverjanje **večkrat (npr. 10x) ponovimo** (torej 10×10=100 izvajanj učnega algoritma) in rezultate povprečimo
- poseben primer prečnega preverjanja je metoda **izloči enega** (angl. leave-one-out, LOO)
  - *k* je enak številu primerov (vsaka testna množica ima samo en primer)
  - najbolj stabilna ocena glede učinkov razbitja na podmnožice
  - časovno zelo zamudno, primerno za manjše množice
- iz meritev na vseh podmnožicah je možno izračunati tudi varianco/ intervale zaupanja



# Pregled




- strojno učenje
  - uvod v strojno učenje
  - učenje odločitvenih dreves
  - učenje dreves iz šumnih podatkov (rezanje dreves)
  - ocenjevanje učenja
  - diskretizacija atributov, obravnava manjkajočih vrednosti
  - naivni Bayesov klasifikator

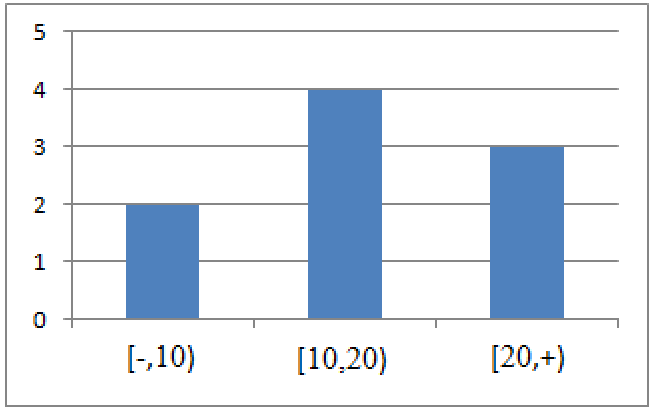
# Različne vrste atributov

- vrste atributov:
  - **diskretni**:
    - **nominalni** ({true, false}, {M, F}, {French, Thai, Burger, Italian})
    - **ordinalni** ({None, Some, Full}, {Low, Med, High})
  - **zvezni/numerični** ($\in \mathbb{R}$)

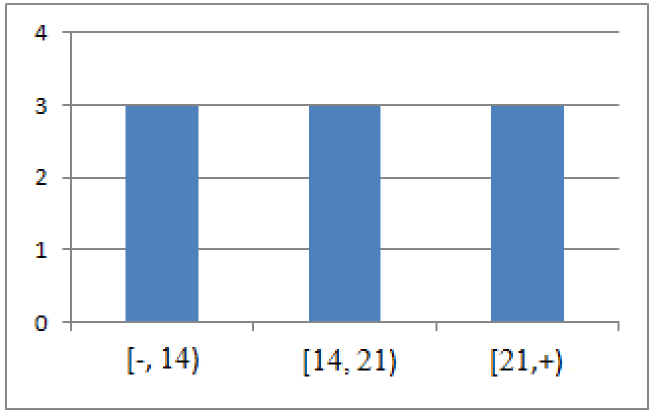
- **Kako obravnavati zvezne/numerične atribute?**
  - Običajno izvedemo diskretizacijo v dva (binarizacija) ali več diskretnih intervalov.
  - Različni pristopi k diskretizaciji:
    - intervali **enake širine** (equal-width)
    - intervali z **enako frekvenco** primerov (equal-frequency)
    - intervali, ki **maksimizirajo informacijski prispevek**

- **Data :** 0, 4, 12, 16, 16, 18, 24, 26, 28
- **Equal width**
  - Bin 1: 0, 4 [-,10)
  - Bin 2: 12, 16, 16, 18 [10,20)
  - Bin 3: 24, 26, 28 [20,+)
- **Equal frequency**
  - Bin 1: 0, 4, 12 [-, 14)
  - Bin 2: 16, 16, 18 [14, 21)
  - Bin 3: 24, 26, 28 [21,+)

**Equal width**



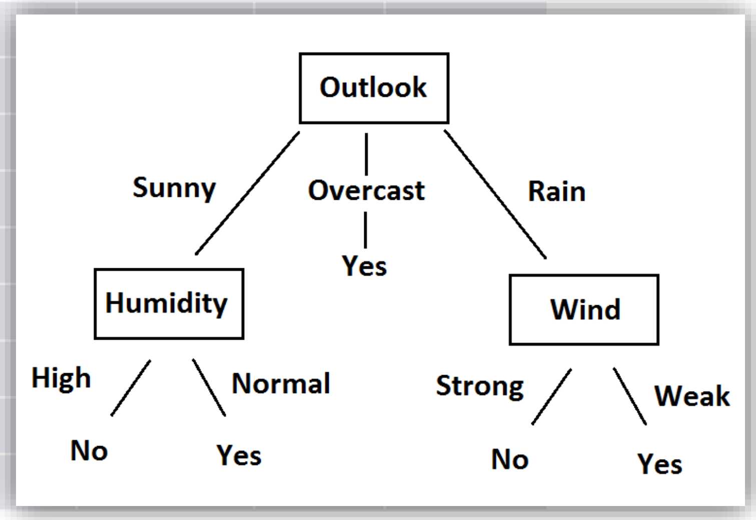

**Equal frequency**



# Diskretizacija z maksimizacijo informacijskega prispevka

- numerični atributi namesto diskretnih?

| Day | Outlook | Temperature | Humidity | Wind | Play |
|---|---|---|---|---|---|
| 1 | Sunny | Hot | High | Weak | No |
| 2 | Sunny | Hot | High | Strong | No |
| 3 | Overcast | Hot | High | Weak | Yes |
| 4 | Rain | Mild | High | Weak | Yes |
| 5 | Rain | Cool | Normal | Weak | Yes |
| 6 | Rain | Cool | | | |
| 7 | Overcast | Cool | | | |
| 8 | Sunny | Mild | | | |
| 9 | Sunny | Cool | | | |
| 10 | Rain | Mild | | | |
| 11 | Sunny | Mild | | | |
| 12 | Overcast | Mild | | | |
| 13 | Overcast | Hot | | | |
| 14 | Rain | Mild | | | |



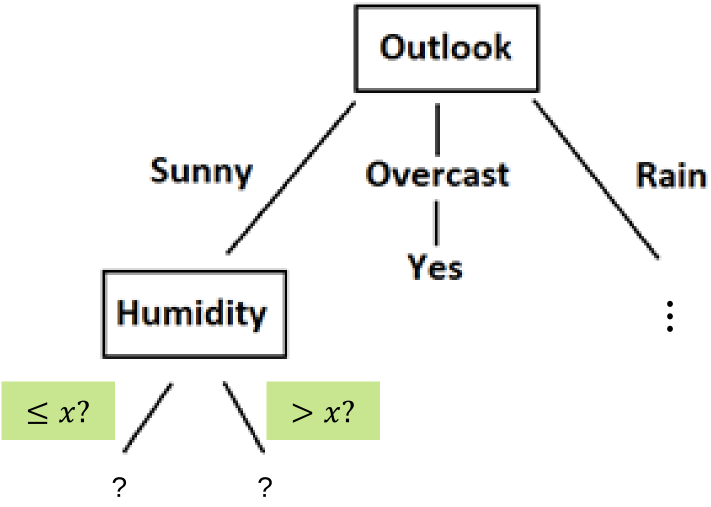

???

| Day | Outlook | Temperature | Humidity | Windy | Play |
|---|---|---|---|---|---|
| 1 | Sunny | 85 | 85 | Weak | No |
| 2 | Sunny | 80 | 90 | Strong | No |
| 3 | Overcast | 83 | 86 | Weak | Yes |
| 4 | Rain | 70 | 96 | Weak | Yes |
| 5 | Rain | 68 | 80 | Weak | Yes |
| 6 | Rain | 65 | 70 | Strong | No |
| 7 | Overcast | 64 | 65 | Strong | Yes |
| 8 | Sunny | 72 | 95 | Weak | No |
| 9 | Sunny | 69 | 70 | Weak | Yes |
| 10 | Rain | 75 | 80 | Weak | Yes |
| 11 | Sunny | 75 | 70 | Strong | Yes |
| 12 | Overcast | 72 | 90 | Strong | Yes |
| 13 | Overcast | 81 | 75 | Weak | Yes |
| 14 | Rain | 71 | 91 | Strong | No |

https://www.slideshare.net/prithwis/lecture02-data-mining-analytics
https://sefiks.com/2017/11/20/a-step-by-step-id3-decision-tree-example/
https://www.chegg.com/homework-help/questions-and-answers/following-table-gives-data-set-deciding-whether-play-cancel-ball-game-depending-weather-co-q27982386

# Diskretizacija z maksimizacijo informacijskega prispevka

- kako diskretizirati atribut Humidity?




| Day | Outlook | Temperature | Humidity | Windy | Play |
|---|---|---|---|---|---|
| 1 | Sunny | 85 | 85 | Weak | No |
| 2 | Sunny | 80 | 90 | Strong | No |
| 8 | Sunny | 72 | 95 | Weak | No |
| 9 | Sunny | 69 | 70 | Weak | Yes |
| 11 | Sunny | 75 | 70 | Strong | Yes |

- primeri možni vrednosti za mejo x: 70, 85, 90
- entropije ob različnih delitvah:
  - $x = 70$ → levo [2,0], desno [0,3]
    $H_{res} = \frac{2}{5}\cdot 0 + \frac{3}{5}\cdot 0 = 0$
  - $x = 85$ → levo [2,1], desno [0,2]
    $H_{res} = \frac{3}{5}\cdot\left(-\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3}\right) + \frac{2}{5}\cdot 0 = 0{,}551$
  - $x = 90$ → levo [2,2], desno [0,1]
    $H_{res} = \frac{4}{5}\cdot 1 + \frac{1}{5}\cdot 0 = 0{,}8$

najnižja residualna entropija (najvišji informacijski prispevek)

- izberemo:
  - LOW: $\leq 70$
  - HIGH: $> 70$

https://www.slideshare.net/prithwis/lecture02-data-mining-analytics
https://sefiks.com/2017/11/20/a-step-by-step-id3-decision-tree-example/

# Obravnava manjkajočih atributov

| Day | Outlook | Temperature | Humidity | Wind | PlayTennis |
|---|---|---|---|---|---|
| 1 | Sunny | Hot | High | Weak | No |
| 2 | Sunny | Hot | High | Strong | No |
| 8 | Sunny | Mild | ??? | Weak | No |
| 9 | Sunny | Cool | Normal | Weak | Yes |
| 11 | Sunny | Mild | Normal | Strong | Yes |

- različni načini **obravnave manjkajočih** atributov, glede na:
  - ali učni algoritem lahko poišče hipotezo **tudi, če manjkajo** vrednosti atributov?
  - ali imamo **dovolj** učnih primerov na razpolago?
  - ali je primerno **nadomeščanje** manjkajočih vrednosti?

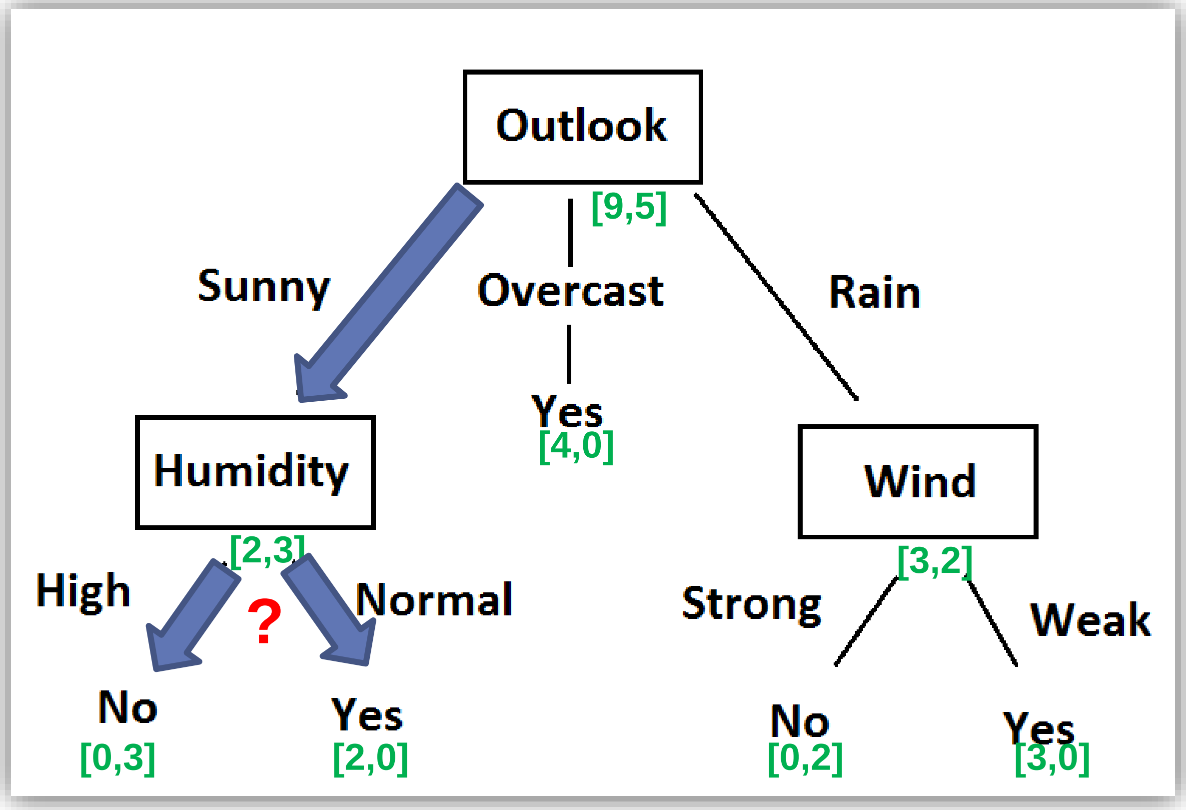
- **pristopi pri učenju**:
  - učimo z manjkajočimi vrednostmi
  - ignoriranje celih učnih primerov z neznanimi vrednostmi
  - uporaba posebne vrednosti NA/UNKNOWN?
  - nadomestiti manjkajočo vrednost (povprečna, najbolj pogosta, najbolj pogosta glede na razred, naključna, napovedana)
  - primer obravnavamo verjetnostno glede na vse možne vrednosti atributa

- **pristopi pri napovedovanju**:
  - verjetnostna klasifikacija glede na vse možne vrednosti atributa (npr. pri drevesu: klasificiramo v več listov)

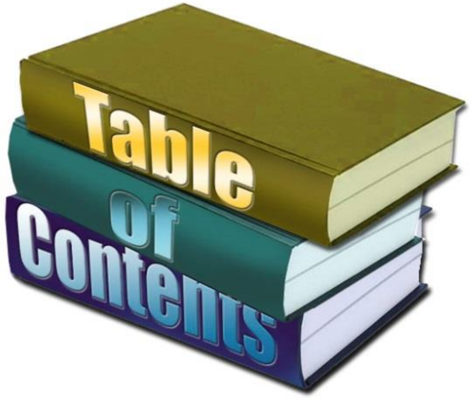
# Obravnava manjkajočih atributov

- Kako klasificirati primer
  Outlook=Sunny,
  Temperature=Mild,
  Humidity=???,
  Wind=Weak ?

- verjetnostna klasifikacija glede na relativno frekvenco:
  - P(razred = No | primer) = 3/5
  - P(razred = Yes | primer) = 2/5



# Pregled



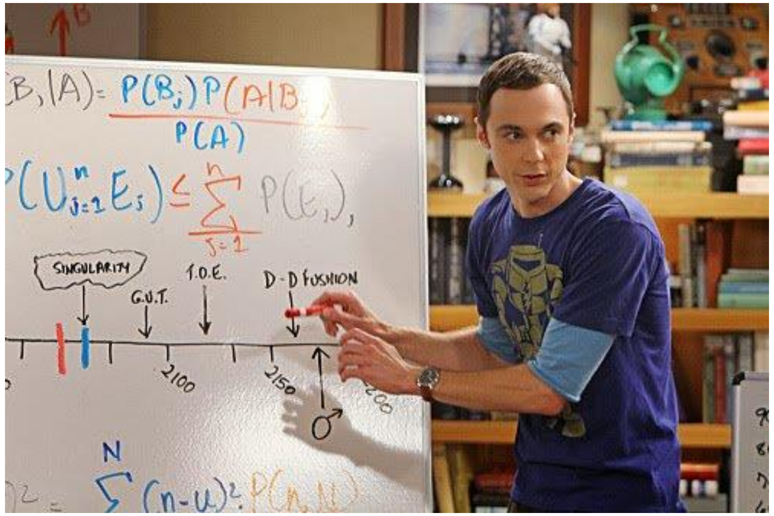

- strojno učenje
  - uvod v strojno učenje
  - učenje odločitvenih dreves
  - učenje dreves iz šumnih podatkov (rezanje dreves)
  - ocenjevanje učenja
  - diskretizacija atributov, obravnava manjkajočih vrednosti
  - naivni Bayesov klasifikator

# Naivni Bayesov klasifikator

- Thomas Bayes, 1702 – 1761
- opomnik iz teorije o verjetnosti:

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

Bayesovo pravilo



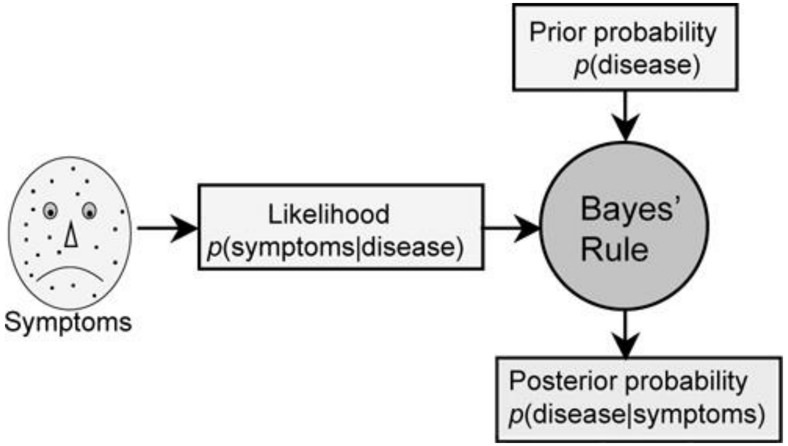
# Naivni Bayesov klasifikator

- zdravniki razpolagajo z **vzročno in statistično informacijo**:
  - verjetnost **izraženih simptomov pri neki bolezni** - $P(opažanje|hipoteza)$
  - verjetnost **določene bolezni** - $P(hipoteza)$
  - verjetnost **določenega simptoma** - $P(opažanje)$

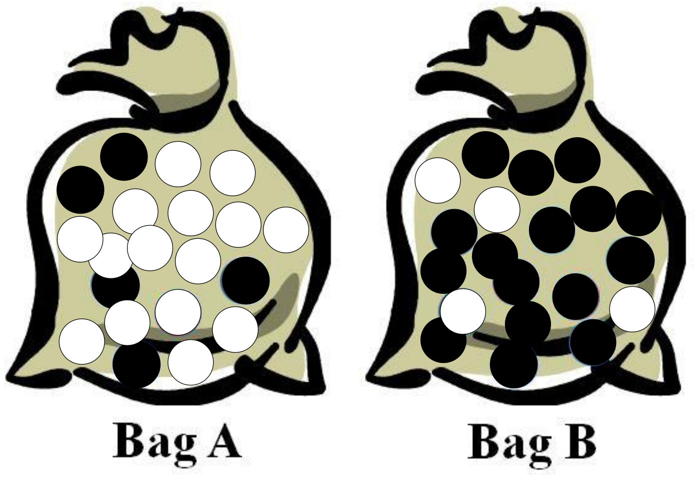
- aplikacija v medicini:

$$P(hipoteza|opažanje) = \frac{P(opažanje|hipoteza) \cdot P(hipoteza)}{P(opažanje)}$$

- Bayesovo pravilo nam izraža **diagnostično pogojno verjetnost** $P(hipoteza|opažanje)$ na podlagi **vzročne pogojne verjetnosti** $P(opažanje|hipoteza)$



# Vaja – opomnik iz verjetnosti




- dve vrsti vrečk s frnikulami:
  - 4 vrečke tipa A (vsaka 5 črnih, 15 belih frnikul)
  - 1 vrečka tipa B (16 črnih, 4 bele frnikule)

- primeri vprašanj, zapisanih kot verjetnosti:
  - Kakšna je verjetnost, da je to vrečka tipa B?
    $P(B) = ?$
  - Kakšna je verjetnost, da naključno izberemo črno frnikulo, če izbiramo iz vrečke tipa B?
    $P(Č|B) = ?$
  - Naključno izberemo eno izmed vrečk in iz nje naključno izberemo frnikulo. Kakšna je verjetnost, da smo izbrali črno frnikulo iz vrečke tipa B?
    $P(BČ) = P(B) \cdot P(Č|B) = ?$
  - Naključno izberemo eno izmed vrečk in iz nje naključno izberemo frnikulo. Kakšna je verjetnost, da smo izbrali črno frnikulo?
    $P(Č) = P(B) \cdot P(Č|B) + P(A) \cdot P(Č|A)$

# Vaja

- Ena vrečka ima poškodovan ovoj tako, da se skozi njega vidi črna frnikula. Kakšna je verjetnost, da je to vrečka tipa B?
  $P(B|Č) =?$
- B = hipoteza, Č = evidenca, opažanje
- verjetnost $P(B|Č)$ lahko določimo iz drugih bolj očitnih verjetnosti z Bayesovo formulo:

$$P(B|Č) = \frac{P(B) \cdot P(Č|B)}{P(Č)}$$

- $P(B) = \frac{1}{5} = 0{,}2$
  $P(Č|B) = \frac{16}{20} = 0{,}8$
  $P(Č) = \frac{4 \cdot 5 + 1 \cdot 16}{5 \cdot 20} = 0{,}360$
- $P(B|Č) = \frac{0{,}2 \cdot 0{,}8}{0{,}360} = 0{,}444$

dve vrsti vrečk s frnikulami:

- 4 vrečke tipa A (vsaka 5 črnih, 15 belih frnikul)
- 1 vrečka tipa B (16 črnih, 4 bele frnikule)

# Naivni Bayes v strojnem učenju

- evidenca → atributi
  hipoteza → razred
- zanima nas, kakšna je verjetnost razreda $C$ pri podanih vrednostih atributov $A_1 = X_1, A_2 = X_2, \dots, A_n = X_n$:

$$P(C|X_1X_2 \dots X_n) = \frac{P(C) \cdot P(X_1X_2 \dots X_n|C)}{P(X_1X_2 \dots X_n)}$$

apriorna verjetnost razreda

Class Prior Probability

Likelihood

$$P(c\,|\,x) = \frac{P(x\,|\,c)\,P(c)}{P(x)}$$

aposteriorna verjetnost napovedi

# Naivni Bayes v strojnem učenju

$$P(hipoteza|evidenca) = \frac{P(hipoteza) \cdot P(evidenca|hipoteza)}{P(evidenca)} \Rightarrow P(C|X_1X_2 \ldots X_n) = \frac{P(C) \cdot P(X_1X_2 \ldots X_n|C)}{P(X_1X_2 \ldots X_n)}$$

- pravilo za produkt verjetnosti: $P(AB) = P(A) \cdot P(B|A)$
- pravilo za produkt verjetnosti s pogojno verjetnostjo: $P(AB|{\color{red}C}) = P(A|{\color{red}C}) \cdot P(B|A{\color{red}C})$

- $$\begin{aligned} P(X_1X_2 \ldots X_n|C) &= P(X_1|C) \cdot P(X_2 \ldots X_n|X_1C) = \\ &= P(X_1|C) \cdot P(X_2|X_1C) \cdot P(X_3 \ldots X_n|X_1X_2C) = \\ &= P(X_1|C) \cdot P(X_2|X_1C) \cdot P(X_3|X_1X_2C) \cdot \ldots \cdot P(X_n|X_1X_2 \ldots X_{n-1}C) \end{aligned}$$
- $P(X_1X_2 \ldots X_n) = P(X_1|X_2 \ldots X_n) \cdot P(X_2|X_3 \ldots X_n) \cdot \ldots \cdot P(X_{n-1}|X_n) \cdot P(X_n)$

- potrebujemo veliko število pogojnih verjetnosti, katerih poznavanje je v praksi težavno
- število kombinacij pogojnih verjetnosti je glede možne vrednosti $X_1X_2 \ldots X_n$ eksponentno
- praktična rešitev: **naivni Bayesov klasifikator**

# Naivni Bayes v strojnem učenju

- predpostavimo, da so atributi med seboj **verjetnostno neodvisni** in poenostavimo:

$$P(X_1X_2 \ldots X_n|C) = P(X_1|C) \cdot P(X_2|X_1C) \cdot \ldots \cdot P(X_n|X_1X_2 \ldots X_{n-1}C)$$
$$P(X_1X_2 \ldots X_n) = P(X_1|X_2 \ldots X_n) \cdot P(X_2|X_3 \ldots X_n) \cdot \ldots \cdot P(X_{n-1}|X_n) \cdot P(X_n)$$

$$\Downarrow$$

$$P(X_1X_2 \ldots X_n|C) \approx P(X_1|C) \cdot P(X_2|C) \cdot \ldots \cdot P(X_n|C)$$
$$P(X_1X_2 \ldots X_n) \approx P(X_1) \cdot P(X_2) \cdot \ldots \cdot P(X_{n-1}) \cdot P(X_n)$$

- približki so dobri, **če so atributi med seboj dovolj neodvisni**
- velja torej:

$$P(C|X_1X_2 \ldots X_n) = \frac{P(C) \cdot P(X_1X_2 \ldots X_n|C)}{P(X_1X_2 \ldots X_n)} \approx \frac{P(C) \cdot \prod_i P(X_i|C)}{\prod_i P(X_i)}$$

konstanten člen, ki je neodvisen od ciljne spremenljivke (če opazujemo samo relativne velikosti napovedi različnih razredov, ga lahko izpustimo)

# Naivni Bayes v strojnem učenju

- Bayesov klasifikator: primer **klasificiramo v razred, ki je najbolj verjeten**:

$$h(X_1 X_2 \ldots X_n) = \operatorname{argmax}_k P(C_k) \cdot \prod_{i=1}^{n} P(X_i|C_k)$$

- **učenje**: ocenimo verjetnosti $P(C_k)$ in $P(X_i|C_k)$ za vse razrede $C_k$ in vrednosti atributov $X_i$
- **napovedovanje**: uporabimo zgornjo enačbo za napovedovanje razreda novim primerom
- *opomba: s poenostavitvijo formule in izpustitvijo imenovalca izgubimo verjetnostno interpretacijo (verjetnosti razredov se ne seštejejo več v 1). Problem rešujemo npr. z normalizacijo rezultatov.*

# Primer

- Zajeli smo podatke za 1000 sadežev, ki so lahko bodisi: *banana*, *pomaranča* ali *drugi sadež* (= vrednosti **razreda**). Za vsakega izmed sadežov smo izmerili, ali je *podolgovat*, *sladek* in *rumen* (= **atributi**). Meritve smo zapisali v tabelo:

| **sadež** | **podolgovat** | | **sladek** | | **rumen** | | **skupaj** |
|---|---|---|---|---|---|---|---|
| | **da** | **ne** | **da** | **ne** | **da** | **ne** | |
| banana | 400 | 100 | 350 | 150 | 450 | 50 | 500 |
| pomaranča | 0 | 300 | 150 | 150 | 300 | 0 | 300 |
| drugo | 100 | 100 | 150 | 50 | 50 | 150 | 200 |
| | 500 | 500 | 650 | 350 | 800 | 200 | 1000 |

- iz tabele lahko razberemo različne verjetnosti, npr.:
  - verjetnosti razredov: $P(banana) = \frac{500}{1000} = 0{,}5$, $P(pomaranča) = 0{,}3$, $P(drugo) = 0{,}2$
  - pogojne verjetnosti: $P(dolg|banana) = \frac{4}{5} = 0{,}8$

    $P(sladek|banana) = 0{,}7$

    $P(rumen|banana) = 0{,}9$

A simple explanation of Naive Bayes Classification, https://stackoverflow.com/questions/10059594/a-simple-explanation-of-naive-bayes-classification

# Primer

| sadež | podolgovat | | sladek | | rumen | | skupaj |
|---|---|---|---|---|---|---|---|
| | da | ne | da | ne | da | ne | |
| banana | 400 | 100 | 350 | 150 | 450 | 50 | 500 |
| pomaranča | 0 | 300 | 150 | 150 | 300 | 0 | 300 |
| drugo | 100 | 100 | 150 | 50 | 50 | 150 | 200 |
| | 500 | 500 | 650 | 350 | 800 | 200 | 1000 |

- Imamo sadež, ki ni podolgovat, ni sladek, je pa rumen. Kateri sadež je to?

$$P(banana|neP, neS, daR)$$
$$\approx P(banana) \cdot P(neP|banana) \cdot P(neS|banana) \cdot P(daR|banana) =$$
$$= \frac{500}{1000} \cdot \frac{100}{500} \cdot \frac{150}{500} \cdot \frac{450}{500} = 0{,}5 \cdot 0{,}2 \cdot 0{,}3 \cdot 0{,}9 = 0{,}027$$

$$P(pomaranča|neP, neS, daR) \approx 0{,}3 \cdot 1 \cdot 0{,}5 \cdot 1 = 0{,}15 \longleftarrow$$

ta sadež je najverjetneje pomaranča

$$P(drugo|neP, neS, daR) \approx 0{,}2 \cdot 0{,}5 \cdot 0{,}25 \cdot 0{,}25 = 0{,}00625$$

# Izpitna naloga

- 2. izpitni rok, 15. 2. 2018 (prilagojena naloga)

## 2. NALOGA (25t):

Podana je učna množica primerov, ki je prikazana v tabeli (*vreme* in *pritisk* sta atributa, *glavobol* pa je razred). Naloge:

**V kateri razred bi naivni Bayesov klasifikator klasificiral učni primer z vrednostmi atributov *vreme=deževno, pritisk=srednji* (verjetnosti računamo z relativno frekvenco)?**

| vreme | pritisk | glavobol |
|---|---|---|
| sončno | nizek | ne |
| sončno | nizek | ne |
| sončno | srednji | da |
| sončno | visok | ne |
| sončno | nizek | ne |
| sončno | nizek | da |
| deževno | srednji | ne |
| deževno | srednji | da |
| deževno | visok | da |

# nomogrami, k najbližjih sosedov, regresija