# ÚVOD DO UMĚLÉ INTELIGENCE
## (CVIČENÍ 10)

Simona Ondrčková

# 7. DOMÁCÍ ÚKOL

Na Marsu je robot, který se má dostat na základnu.

Robot nepřistál přímo na základně musí tam dojet. Má rozbité motory a často jede jinam jako chce.

Funguje mu ale lokalizace, takže ví kde je.

Každý krok ho stojí energii. Pro každou pozici je známo kolik energie stojí jí projít.

Najděte nejkratší cestu na základnu.

- Nelze předem spočítat optimální cestu
- V každé pozici počítejte nejlepší pokyn, který může robot zvolit.

Přesné zadání: https://gitlab.mff.cuni.cz/finkj1am/introai

# 8. DOMÁCÍ ÚKOL

2 hráči, kteří odebírají kameny očíslované 1,2,..,n.

Hráč může vzít jakýkoliv kámen, jež je násobkem nebo dělitelem předchozího odebraného čísla.

Pokud hráč nemůže odebrat kámen prohrál.

Napište funkci, která dostane seznam zbývajících kamenů a poslední odebraný kámen a určí zda je situace vyhrávající a vrátí jaký kámen se má odebrat v dalším tahu.

# STROJOVÉ UČENÍ

Jaké znáte příklady strojové učení?

Proč nepoužíváme neuronové sítě na všechno?

Přístupy strojového učení:

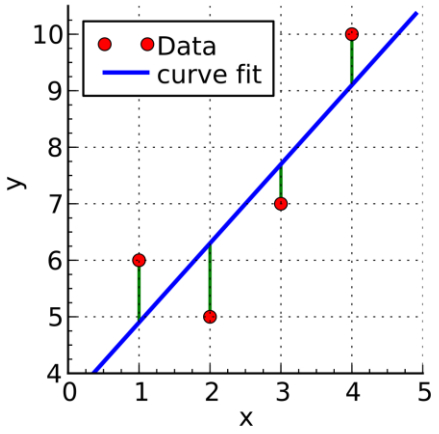
Učení s učitelem (supervised learning)

Učení bez učitele (unsupervised learning)

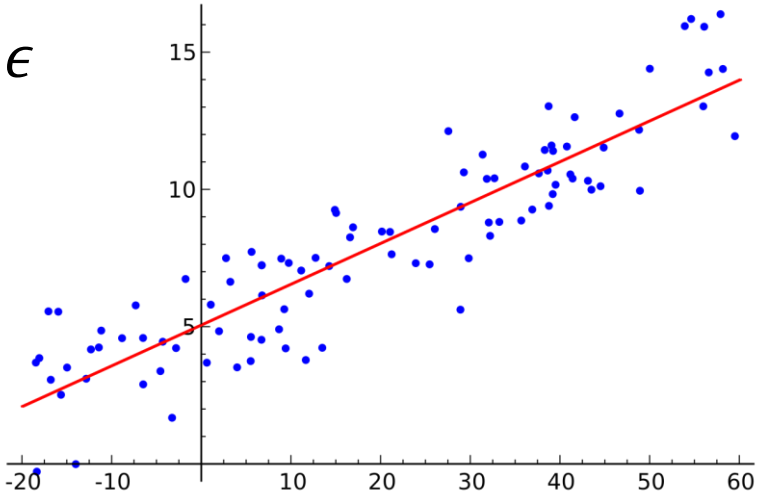
Zpětnovazebné učení (reinforcement learning)

# REGRESE

**Lineární regrese:**




$$y_i = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n + \epsilon$$




**Polynomiální regrese...**

# KLASIFIKACE

Rozdělit data do více skupin:

Například je pacient nemocný?  Co to je za objekt? Je to spam nebo ne?

| | Spam | Běžný email |
|---|---|---|
| Označeno za spam | True positive | False positive |
| Označeno za běžný email | False negative | True negative |

Accuracy: $\frac{TP+TN}{TP+FP+TN+FN}$

Precision: $\frac{TP}{TP+FP}$, Recall: $\frac{TP}{TP+FN}$

# KLASIFIKACE

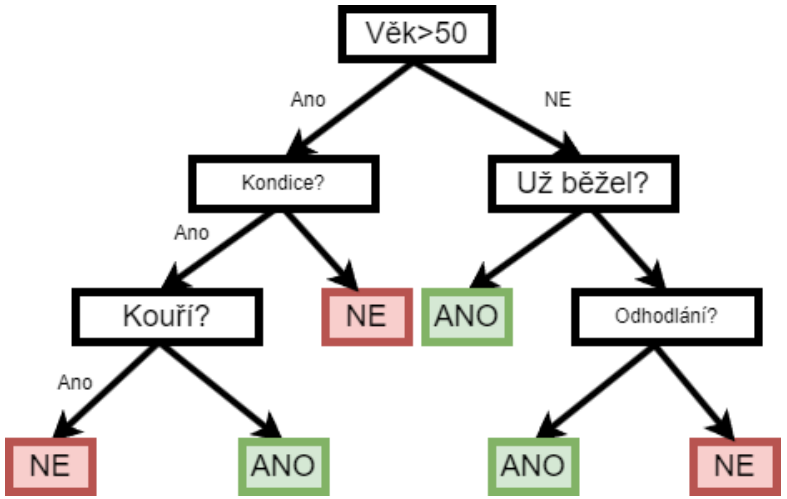
Proč máme tolik způsobů vyhodnocení?

Určete kvalitu testů pokud 10% emailů jsou spamy a

1) Test je vždy negativní
2) Test ze spamů pozná 10% a jinak je negativní
3) Test je vždy pozitivní

Accuracy: $\frac{TP+TN}{TP+FP+TN+FN}$, Precision: $\frac{TP}{TP+FP}$, Recall: $\frac{TP}{TP+FN}$

# ROZHODOVACÍ STROMY




Rozhodněte zda uběhnou maraton:

Julie: věk 20, nekouří, neběžela, v dobré kondici, není odhodlaná.

Honza:  věk 80, kouří, v dobré kondici, neběžel, je odhodlaný

Josef: věk 49, už běžel, kouří, v dobré kondici, není odhodlaný.

# ROZHODOVACÍ STROMY

**Jak vytvořit rozhodovací strom?**

Vyberte atribut podle kterého strom rozvětvíte

Rozdělte vzorky podle kritéria

Pokračujte rekurzí dokud máte vzorky stejné kategorie.

**Jak vybrat atribut podle kterého rozdělovat?**

Podle entropie. $n_i$je počet vzorků v kategorii i. Potom $p_i = \frac{n_i}{n}$

Entropie: $H = -\sum_i p_i log_2(p_i)$

Hledáme atribut co nejvíce sníží entropii. Atribut rozdělí vzorky do skupiny velikosti k1 s entropií H1 a skupiny velikosti k2 s entropií H2.

$$\frac{k_1}{k}H1 + \frac{k_2}{k}H1$$

# ROZHODOVACÍ STROMY

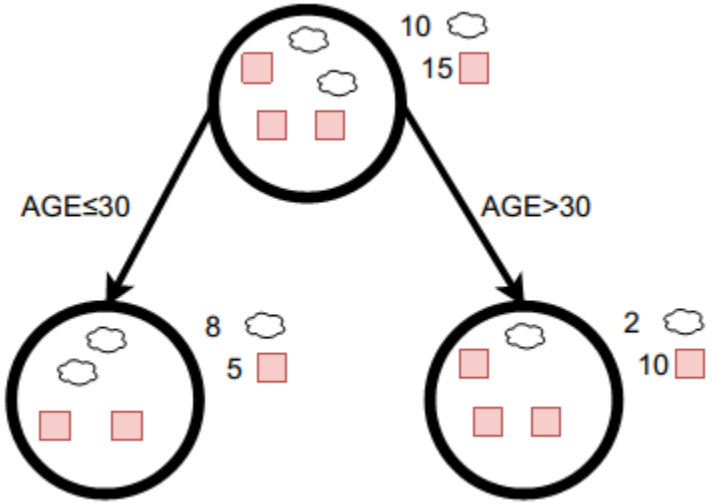
1) Vypočítejte pravděpodobnosti jednotlivých

tvarů ve všech uzlech.

2) Vypočítejte entropii ve všech uzlech (rodič, (age ≤30,age>30).

3) Vypočítejte průměrnou entropii přes oba věkové uzly.

4) Vypočítejte zisk informace (information gain).



# ROZHODOVACÍ STROMY

**o=oblak, c=čtverec, R=rodič, L=levý uzel, P=pravý uzel.**

**1)** **R:** $p(o) = \frac{10}{25} = 40\%, p(c) = \frac{15}{25} = 60\%$

**L:** $p(o) = \frac{8}{13} \cong 61{,}5\%, p(c) = \frac{5}{13} \cong 38{,}5\%,$

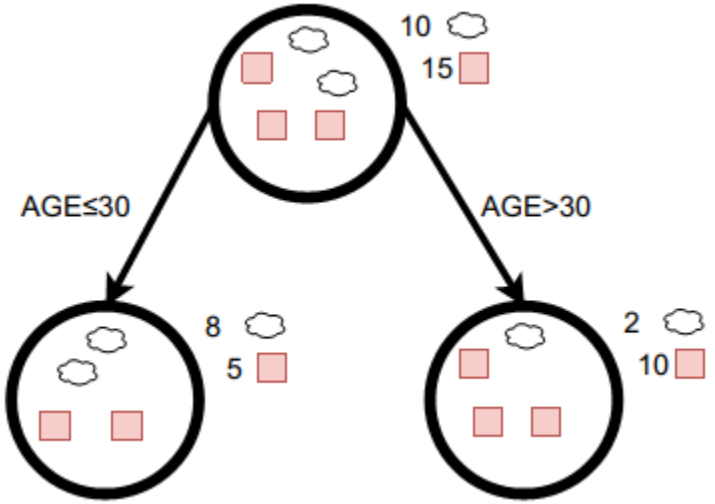
**R:** $p(o) = \frac{2}{12} \cong 16{,}7\%, p(c) = \frac{2}{12} \cong 83{,}3\%,$

**2) Entropie rodiče:**

$$H(R) = -\sum_i p_i \log_2(p_i) = -\frac{10}{25} log \frac{10}{25} - \frac{15}{25} log \frac{15}{25} \cong 0{,}97$$

$$H(L) = -\frac{8}{13} log \frac{8}{13} - \frac{5}{13} log \frac{5}{13} \cong 0{,}961$$

$$H(P) = -\frac{2}{12} log \frac{2}{12} - \frac{10}{12} log \frac{10}{12} \cong 0{,}65$$



# ROZHODOVACÍ STROMY

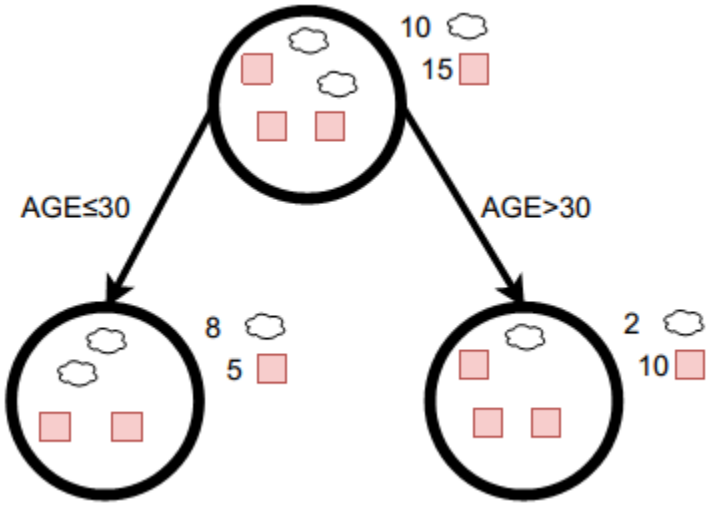
**Entropie:** $H(R) \cong 0{,}97, H(L) \cong 0{,}961, H(P) \cong 0{,}65$

**3) Vypočítejte průměrnou entropii přes oba věkové uzly:**

$$H(vek) = \frac{13}{25} * 0{,}961 + \frac{12}{25} * 0{,}65 \cong 0{,}81$$

**Vypočítejte zisk informace (information gain).**

**4)** $Gain(rodic, vek) = 0{,}97 - 0{,}81 = 0{,}16$



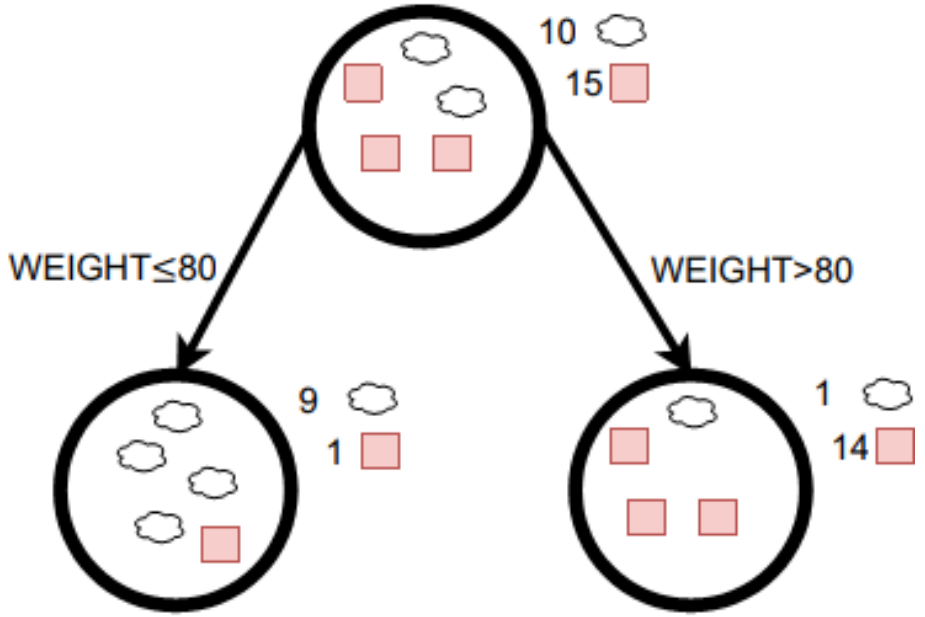
# ROZHODOVACÍ STROMY

1) Vypočítejte pravděpodobnosti jednotlivých tvarů ve všech uzlech.

2) Vypočítejte entropii v uzlech váhy

3) Vypočítejte průměrnou entropii přes oba váhové uzly.

4) Vypočítejte zisk informace (information gain).



# ROZHODOVACÍ STROMY

**R:** $p(o) = \frac{10}{25} = 40\%, p(c) = \frac{15}{25} = 60\%, H(R) \cong 0{,}97$

**L:** $p(o) = \frac{9}{10} = 90\%, p(c) = \frac{1}{10} = 10\%,$

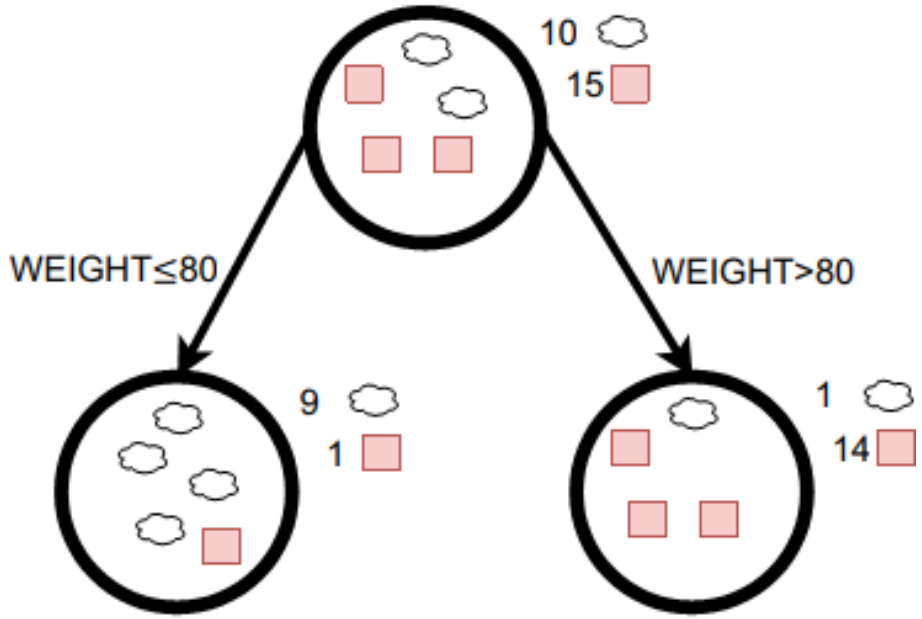
**P:** $p(o) = \frac{1}{15} = 6{,}7\%, p(c) = \frac{14}{15} = 93{,}3\%,$

$H(L) = -\frac{9}{10} log \frac{9}{10} - \frac{1}{10} log \frac{1}{10} \cong 0{,}468$

$H(P) = -\frac{1}{15} log \frac{1}{15} - \frac{14}{15} log \frac{14}{15} \cong 0{,}353$

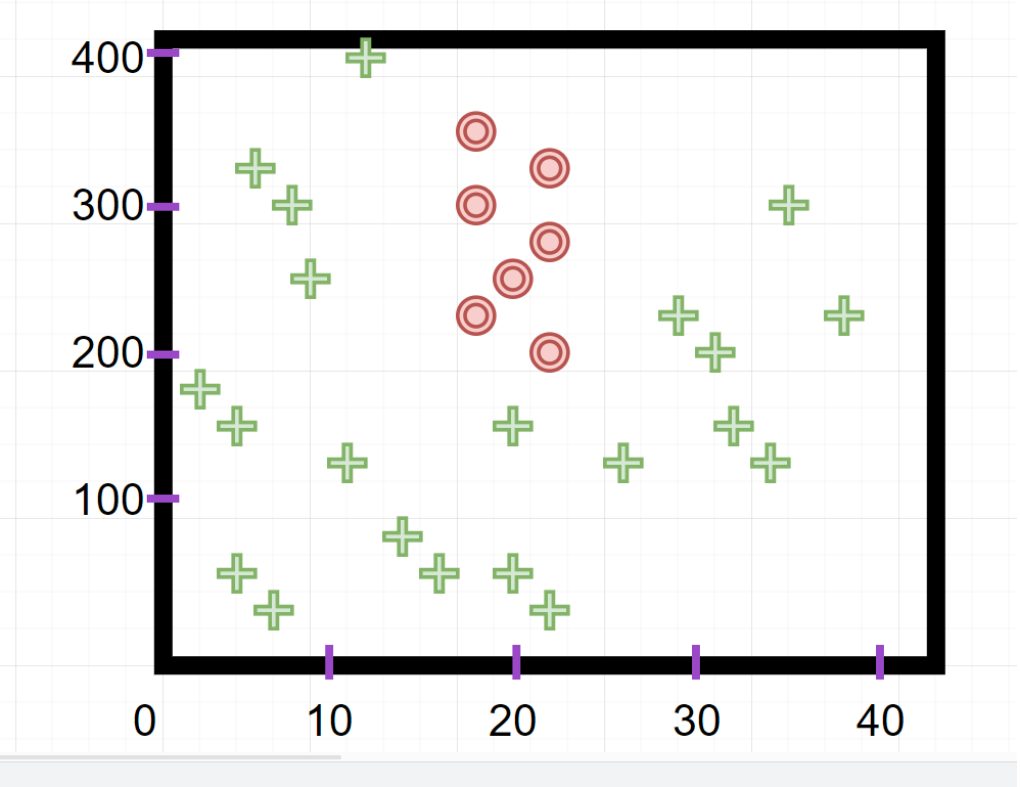
$H(vaha) = \frac{10}{25} * 0{,}468 + \frac{15}{25} * 0{,}353 \cong 0{,}399$

$Gain(rodic, vaha) = 0{,}97 - 0{,}399 = 0{,}58$



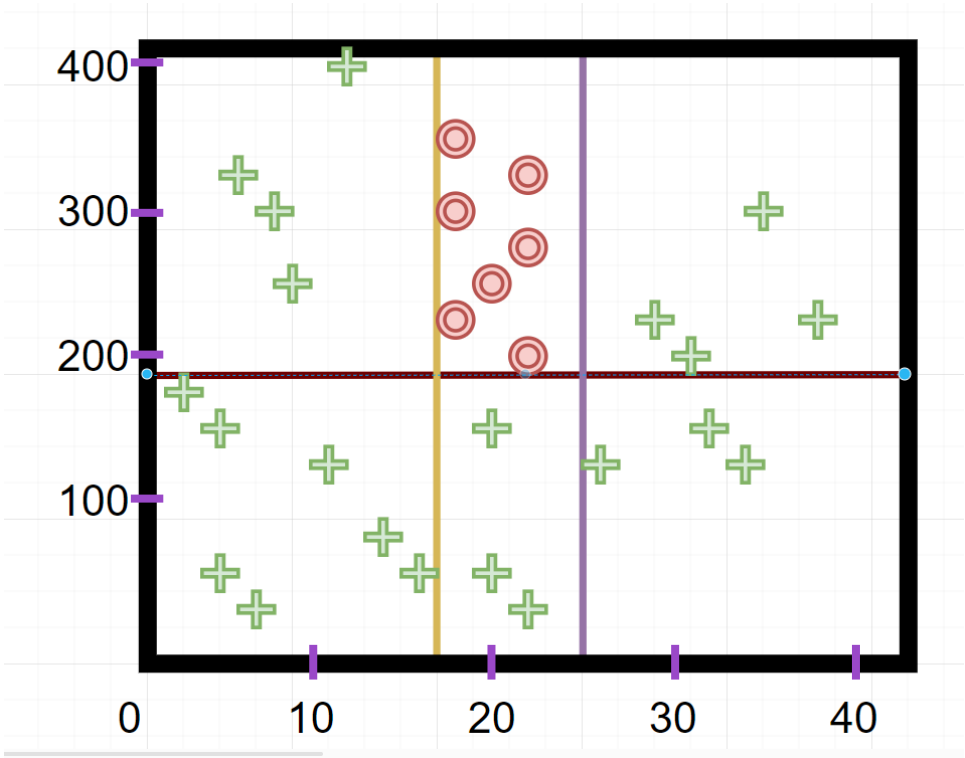
# ROZHODOVACÍ STROM

Vytvořte rozhodovací strom  pro tento problém:



# ROZHODOVACÍ STROM

Podle tohoto stromu určete kategorii vzorkům: (30,210) a (21,390).



# 9. DOMACÍ ÚKOL

Chceme rozpoznat, zda má žena cukrovku.

Máme k dispozici soubor s diagnostikami lidí včetně cukrovky k učení i testování.

Nastavte parametry rozhodovacího stromu, tak aby byla co největší úspěšnost.

Výsledek musí být statisticky relevantní

Odevzdáte zprávu v PDF:

Popíšete parametry co ovlivnily výsledek

Napište nejlepší výsledky a parametry, kterých jste dosáhli.

Nakreslete nejlepší rozhodovací strom.

Vytvořte graf závislosti a úplnosti na velikosti testovací množiny.