AI Akademie

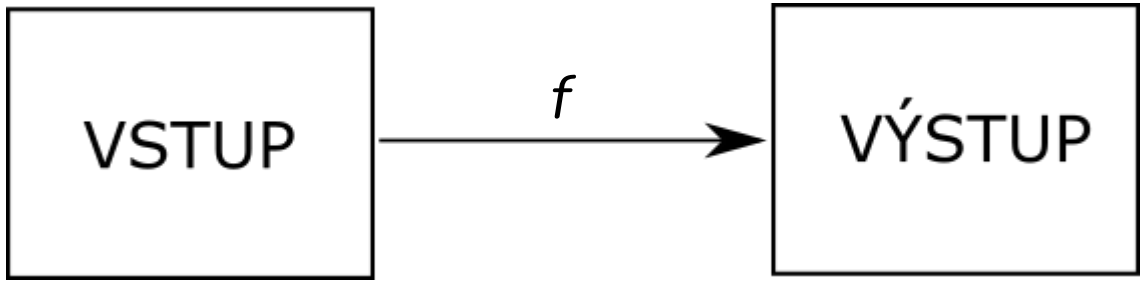
Kapitola 7: Klasifikace

Machine Learning College

prg.ai

PRAHA
PRAGUE
PRAGA
PRAG

Strojové učení



Úlohou strojového učení je na základě příkladů vstupů a výstupů nalézt funkci *f*, která pro nový vstup určí odpovídající výstup.

Příklady dvojic vstupů a výstupů nazýváme *trénovací data*.

V současnosti je to nejrozšířenější metoda umělé inteligence s největšími dopady.

Machine Learning College

Strojové učení - příklady

$\xrightarrow{f}$ pes

klasifikace obrázků

hello $\xrightarrow{f}$ ahoj

strojový překlad AJ -> ČJ

90 km/h $\xrightarrow{f}$ 5,1 l

Predikce spotřeby auta podle průměrné rychlosti

Machine Learning College

Regrese vs. klasifikace

Klasifikace - výstupem je nějaká kategorie (třída). Například *barva*, *binární hodnota (ano, ne)*, *den v týdnu*, *typ auta apod.*

Regrese -výstupem je číselná hodnota. Například *cena*, *teplota*, *počet lidí v místnosti apod.*

Machine Learning College

Klasifikace - příklad

Rozlišení jablek a hrušek

vstup

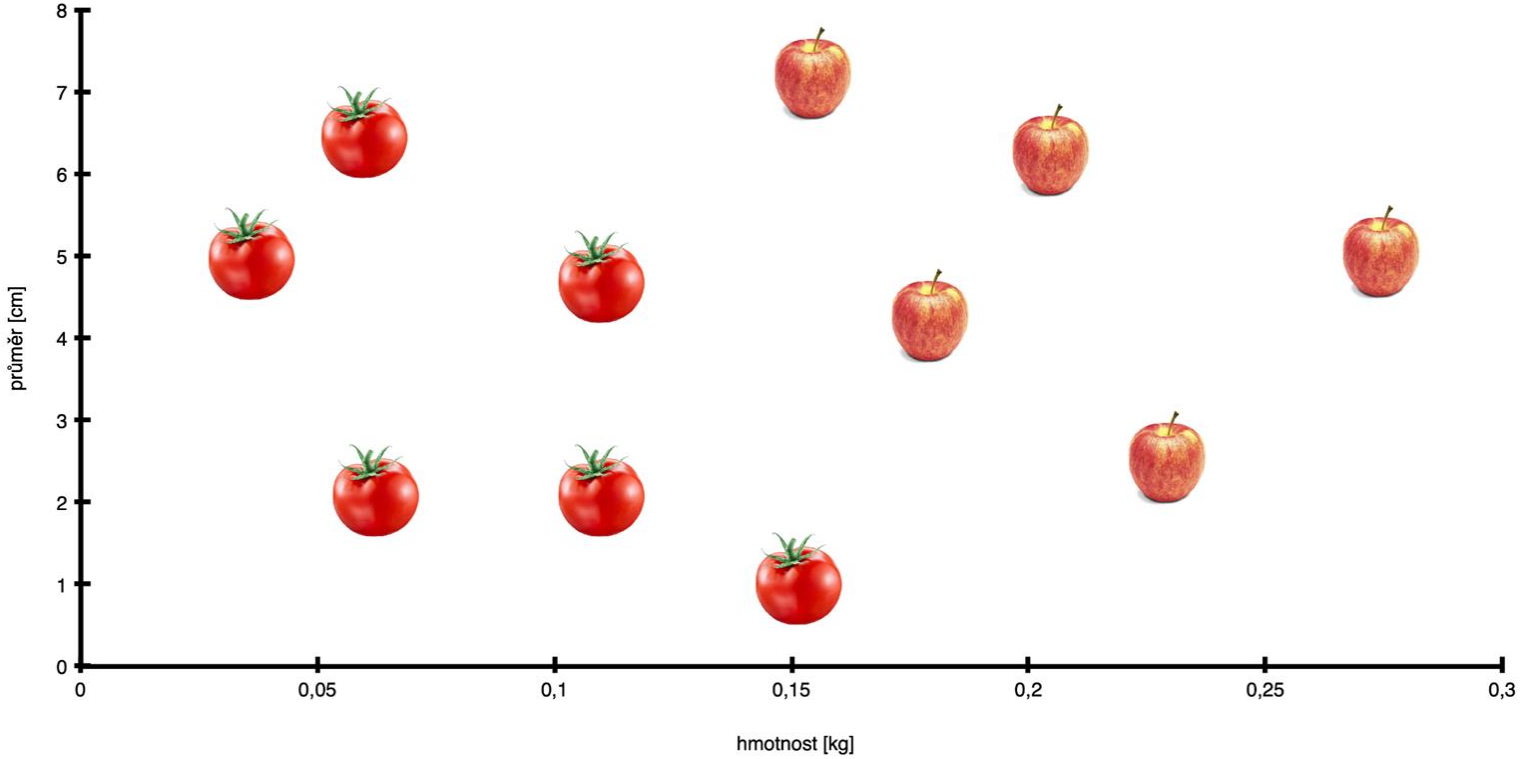
tvar	barva	hmotnost (g)
kulatý	červená	146
šišatý	žlutá	120
šišatý	zelená	187
kulatý	červená	155

výstup

druh ovoce
jablko
hruška
hruška
jablko

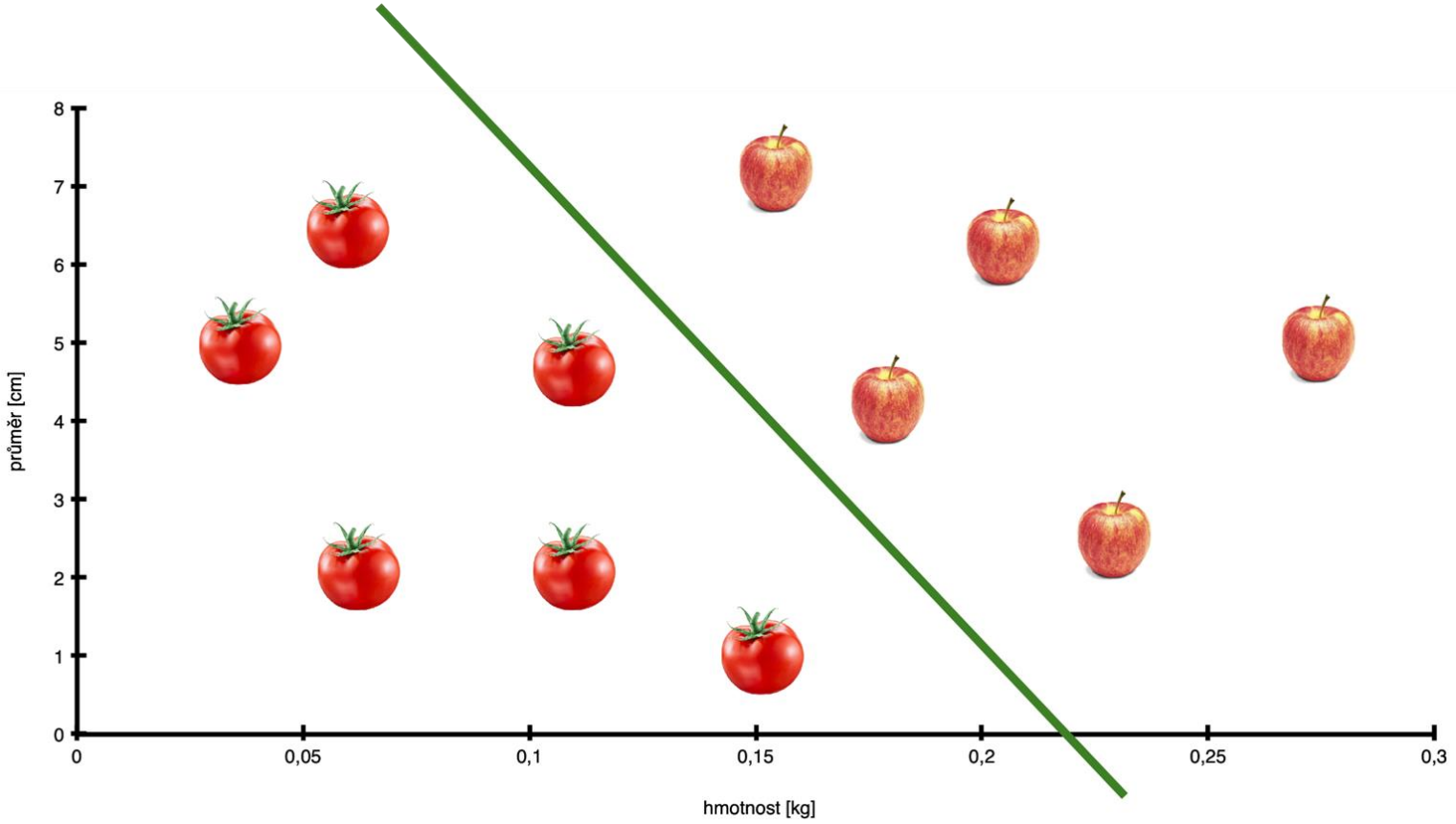
Machine Learning College

Klasifikace - příklad



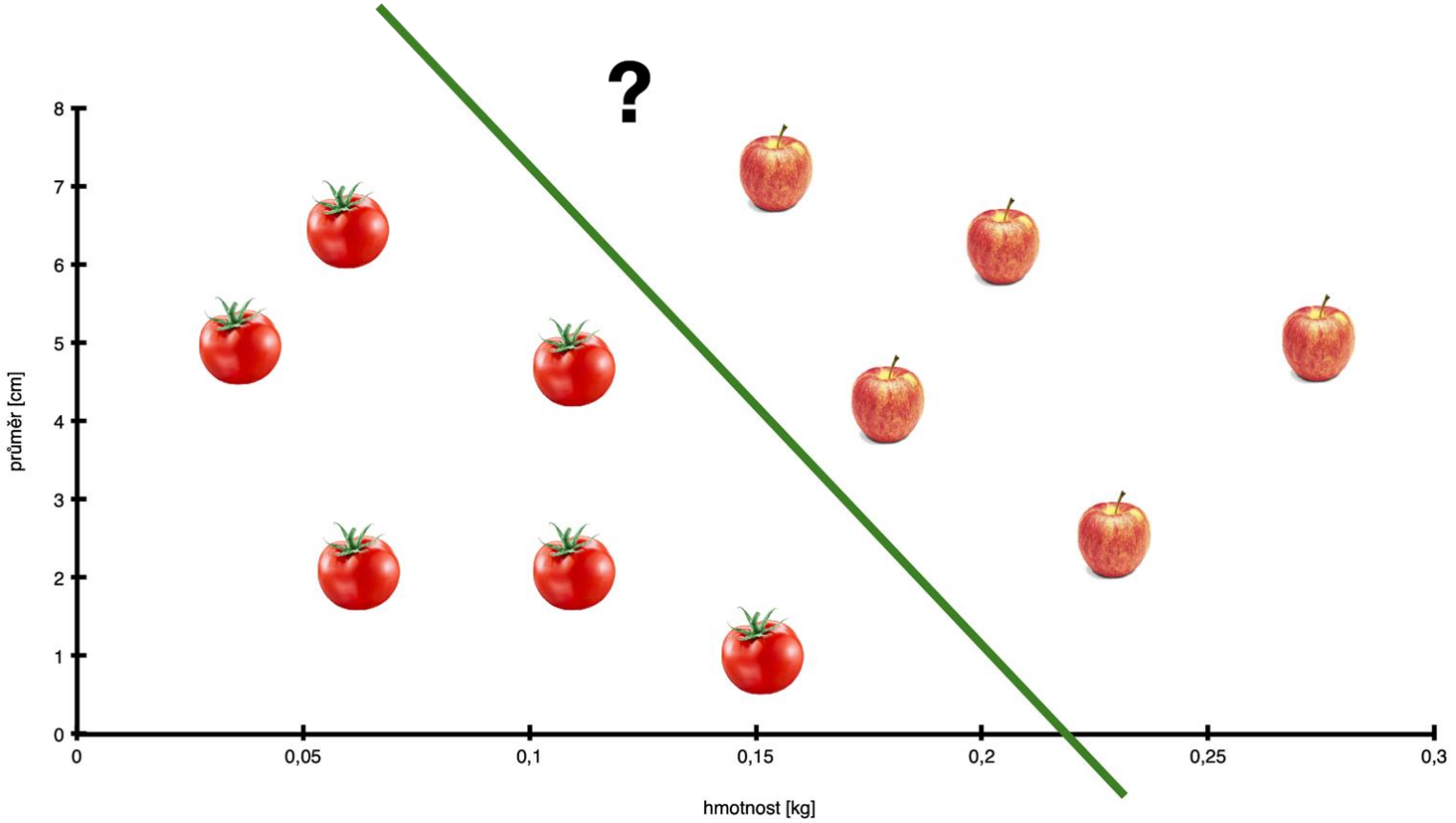
Machine Learning College

Klasifikace - příklad



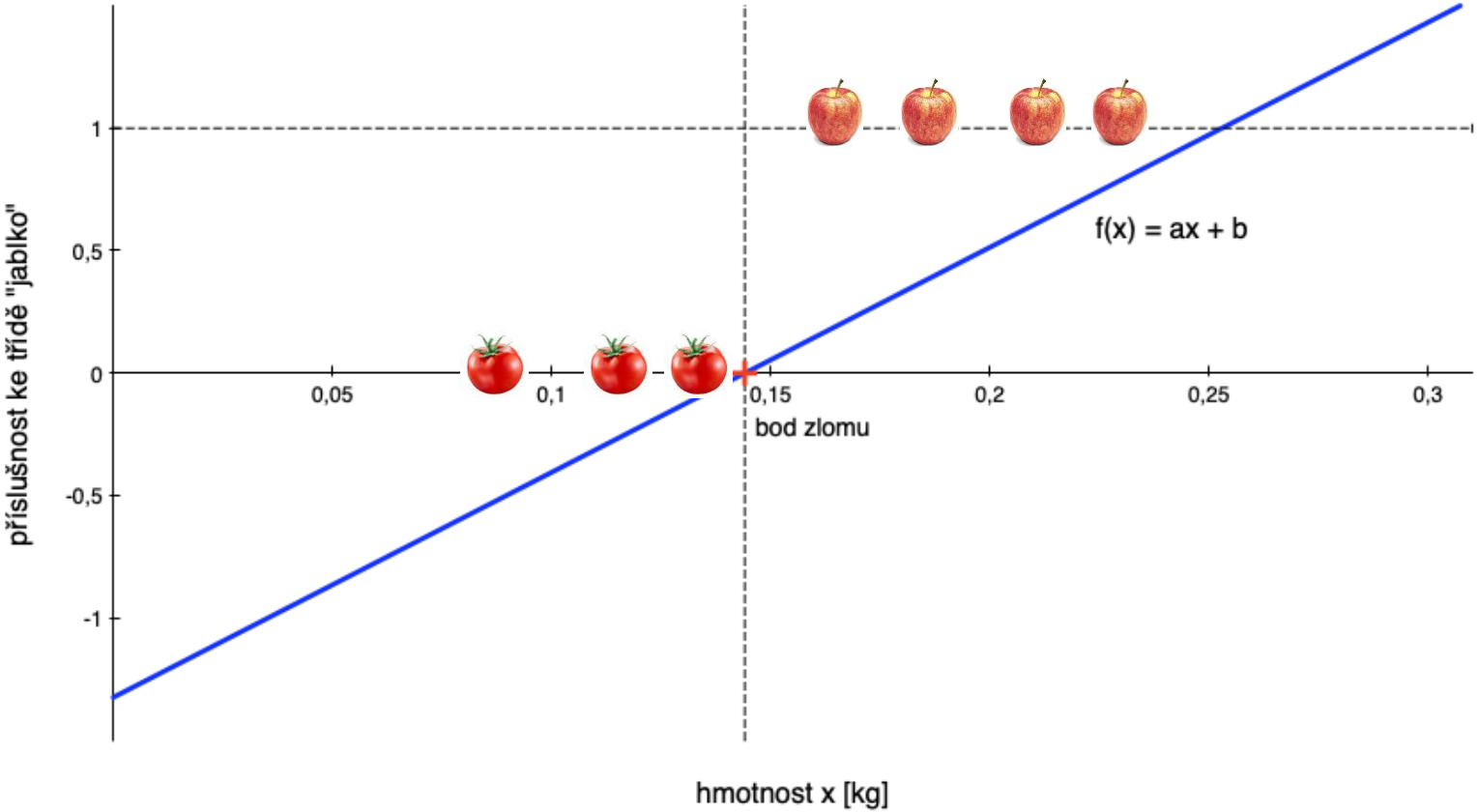
Machine Learning College

Klasifikace - příklad



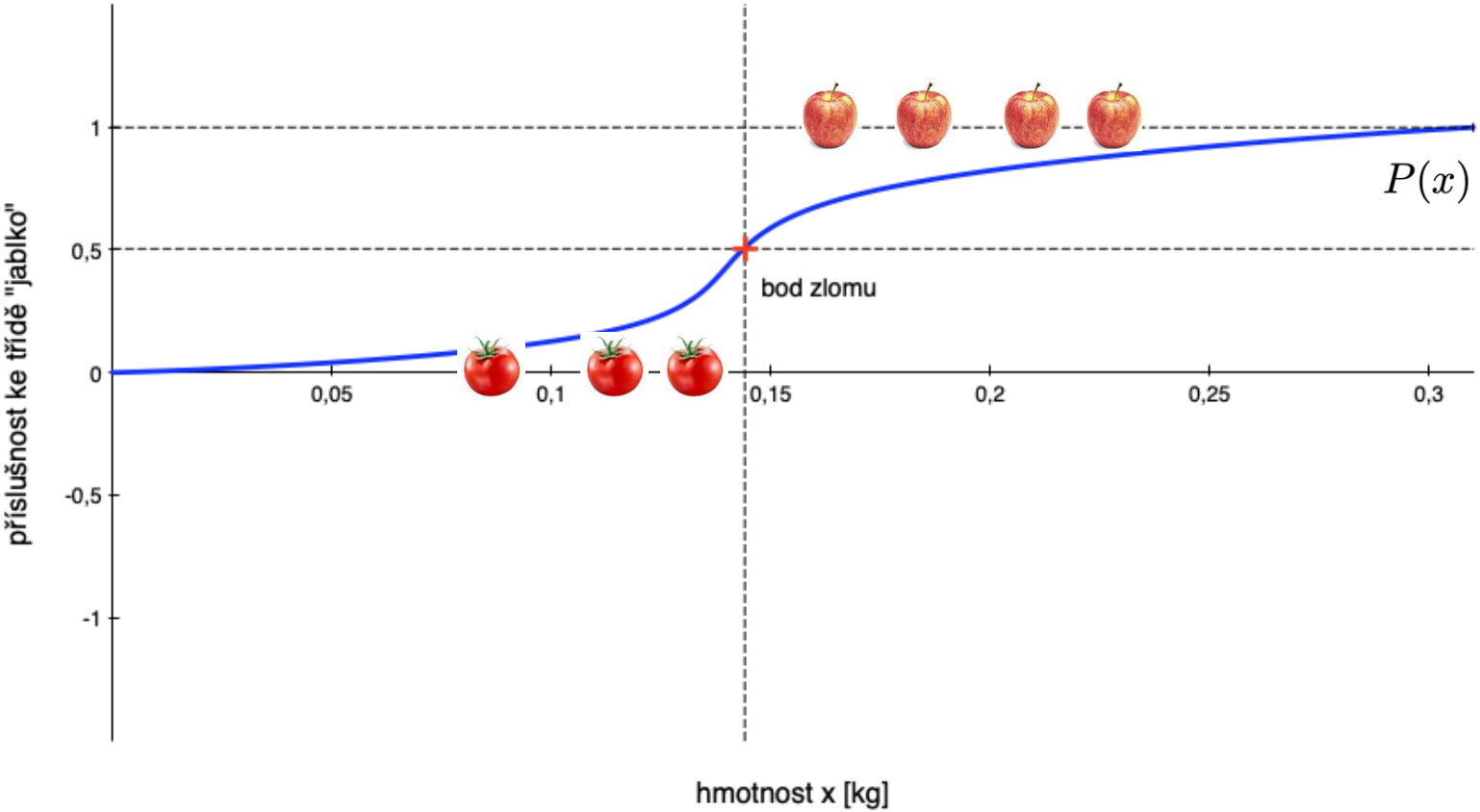
Machine Learning College

Logistická regrese



Machine Learning College

Logistická regrese



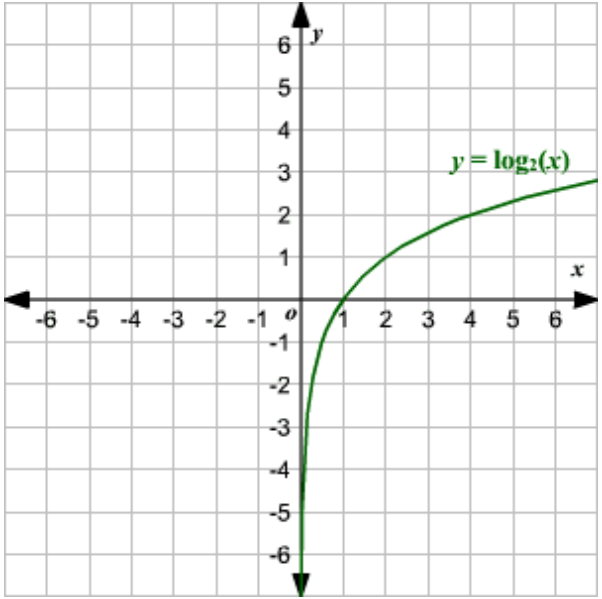
$$P(x) = \frac{1}{1 + e^{-(ax+b)}}$$

Machine Learning College

Logistická regrese - křížová entropie

hodnota trénovacích dat (y)	predikce (pr)	entropie (e)
0 (rajče)	0.1	0.152
1 (jablko)	0.9	0.152
1 (jablko)	0.23	2.12
0 (rajče)	0.99	6.644

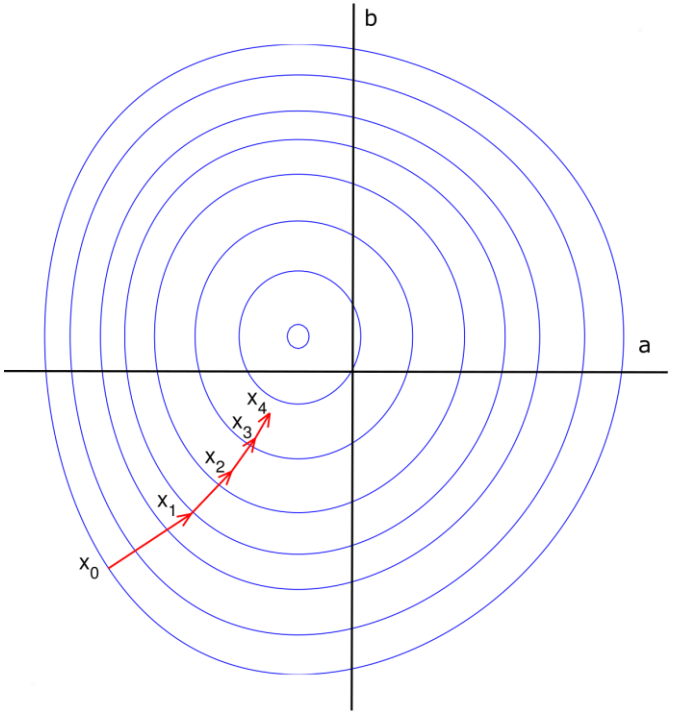
pokud y = 1, pak **e = $-\log_2(pr)$**
pokud y = 0, pak **e = $-\log_2(1-pr)$**



Machine Learning College

Logistická regrese - trénování

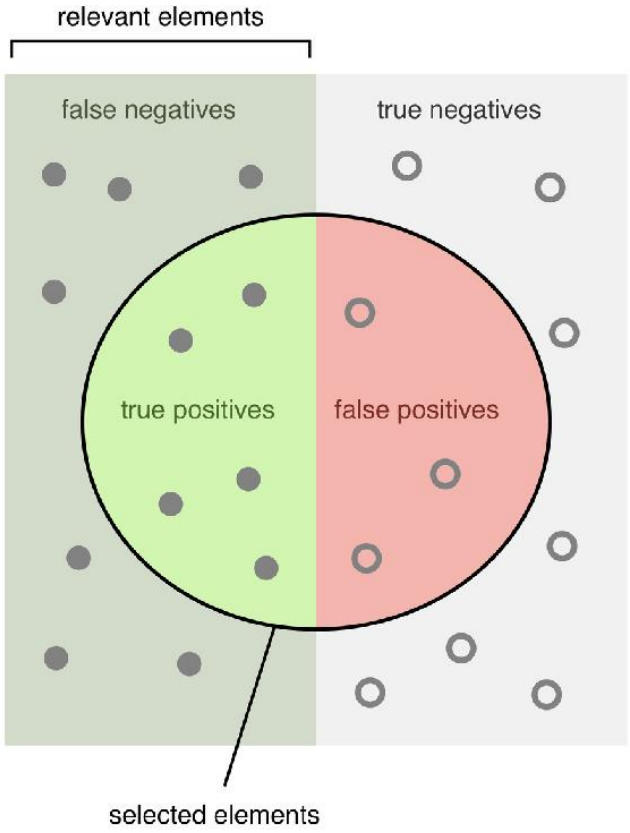
hodnota trénovacích dat (y)	predikce (pr)	entropie (e)
0 (rajče)	0.1	0.152
1 (jablko)	0.9	0.152
1 (jablko)	0.23	2.12
0 (rajče)	0.99	6.644



Celková chyba, kterou minimalizujeme = průměrná entropie všech trénovacích příkladů

Machine Learning College

Klasifikace - vyhodnocení kvality



Správnost $$\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn}$$

Přesnost $$\text{Precision} = \frac{tp}{tp + fp}$$

Pokrytí $$\text{Recall} = \frac{tp}{tp + fn}$$

F-míra $$\text{F} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Machine Learning College

Klasifikace do více než dvou tříd

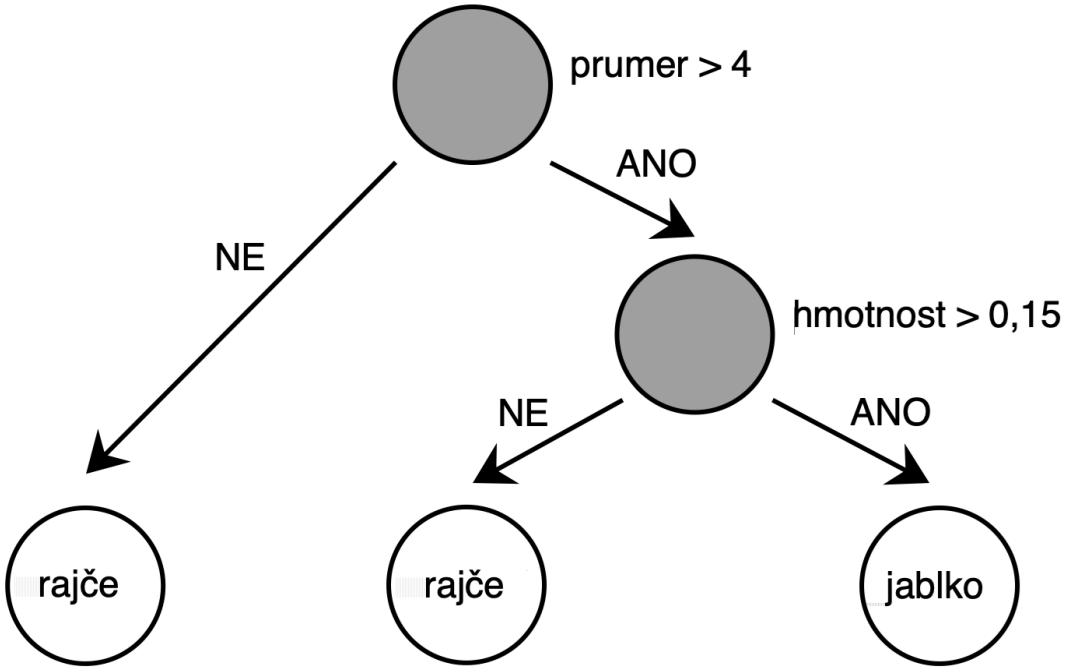
Existuje rozšíření logistické regrese, které umí pracovat rovnou s více než dvěma třídami.

Pokud máme pouze binární klasifikátor (klasifikátor do 2 tříd), lze ho použít pro klasifikaci do *k* tříd jednou z následujících dvou strategií:

a) **jeden versus ostatní** - pro každou z *k* tříd vytvoříme jeden binární klasifikátor, který bude rozlišovat danou třídu od všech ostatních. Zvítězí ta třída, jejíž klasifikátor si bude nejjistější
b) **jeden versus jeden** - vytvoříme *k* * (*k*-1) / 2 klasifikátorů pro všechny možné dvojice tříd. Výsledná třída je poté daná hlasováním všech klasifikátorů.

Machine Learning College

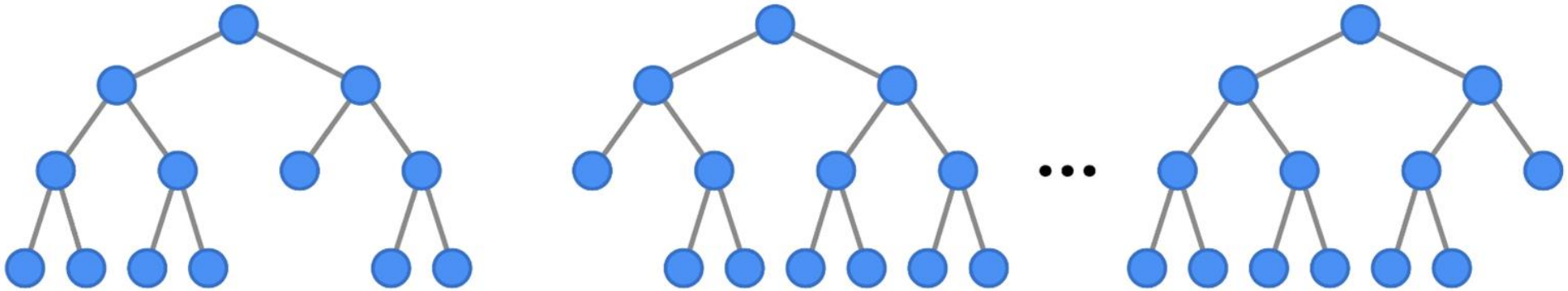
Rozhodovací strom pro klasifikaci



- Při trénování hledáme takový binární strom dané hloubky, který bude mít minimální chybu na trénovacích datech.
- V uzlech může být libovolná podmínka.
- Predikce je uložena v koncových uzlech (listech).

Machine Learning College

Více rozhodovacích stromů - les



- Náhodný les kombinuje více rozhodovacích stromů “hlasováním”. Nejčastější predikce zvítězí.
- Jednou z nejběžnějších implementací je Random forest (náhodný les).
- U random forest je každý strom je vytvořen z náhodně vybrané podmnožiny trénovacích dat, proto je každý strom jiný.

Machine Learning College