AI Akademie

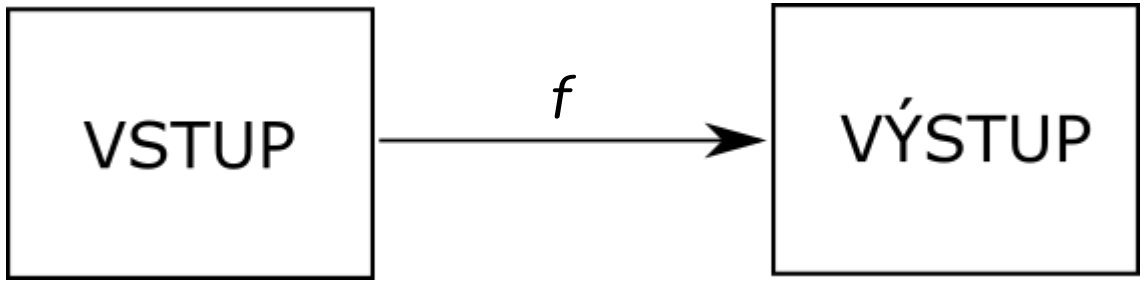
Kapitola 6: Regrese

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

Regrese - příklad

Predikce ceny bytu

vstup

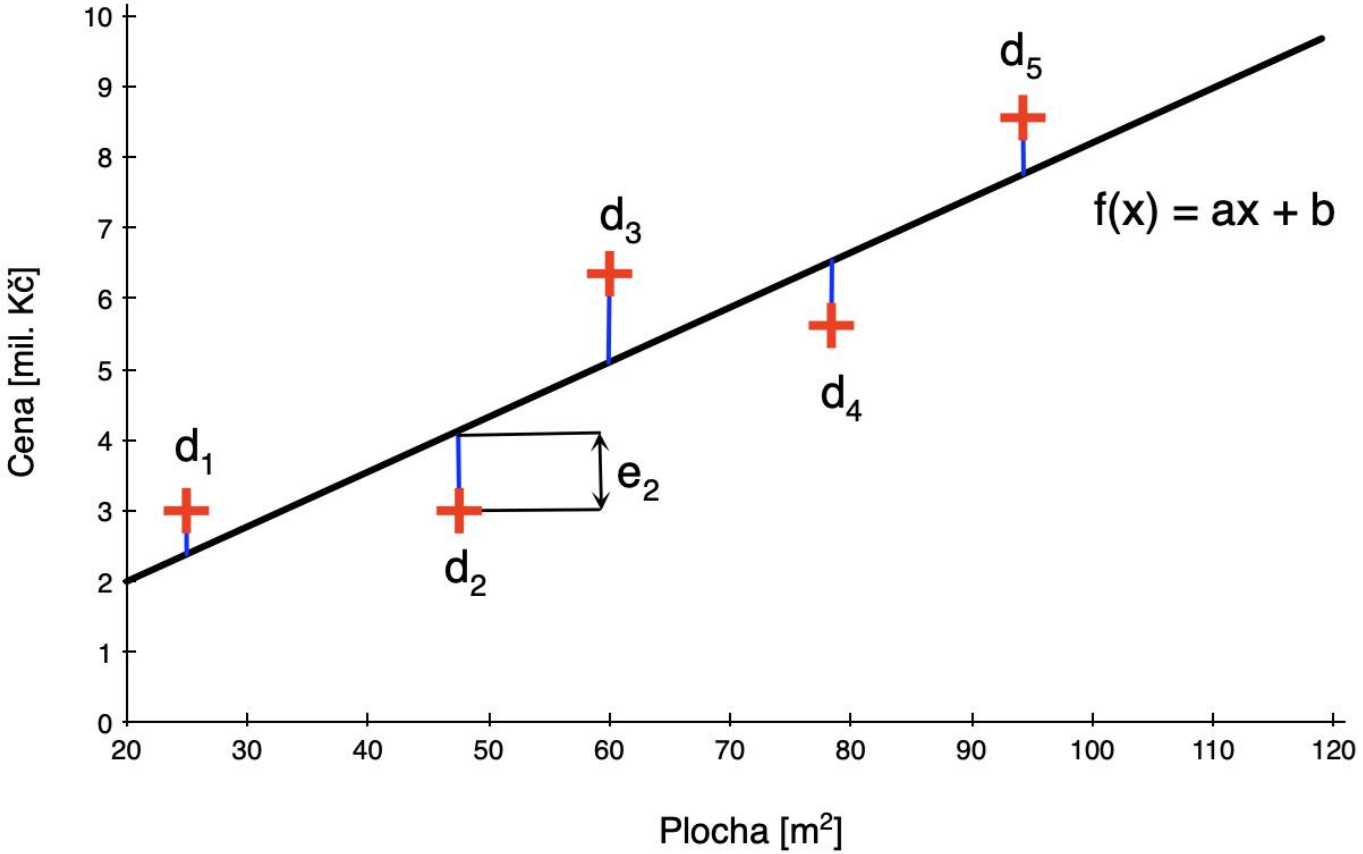
plocha	patro	počet místností
42	7	2
105	3	3
67	1	2
224	3	4

výstup

cena (mil. Kč)
3,2
6,8
4,1
13,9

Machine Learning College

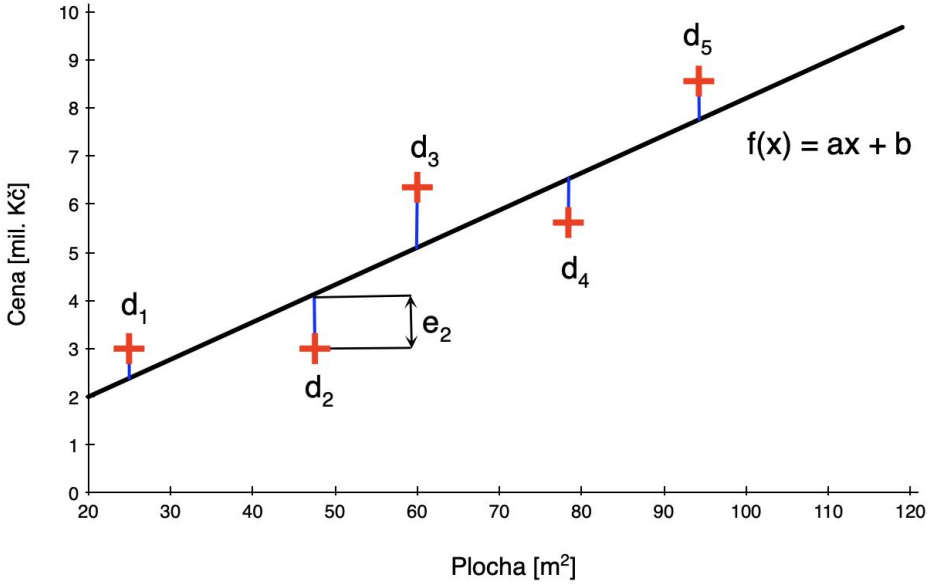
Lineární regrese



Machine Learning College

Lineární regrese

$d_1..d_5$ - **trénovací data**
f(x) - **model**
e_2 - **chyba predikce pro dům č. 2**
MAE - **průměrná absolutní chyba (Mean Absolute Error)**
MSE - **průměrná kvadratická chyba (Mean Squared Error)**

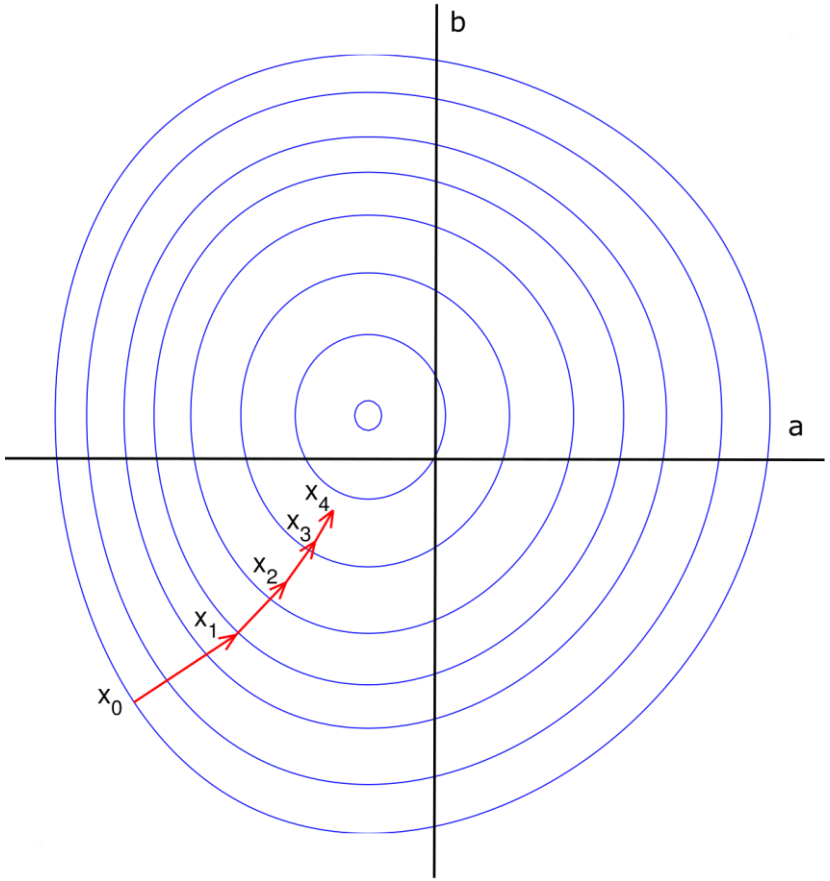


$$MSE = \frac{e_1^2 + e_2^2 + e_3^2 + e_4^2 + e_5^2}{5}$$

$$MAE = \frac{|e_1| + |e_2| + |e_3| + |e_4| + |e_5|}{5}$$

Machine Learning College

Lineární regrese - trénování



Trénování modelu spočívá v hledání parametrů a, b tak, aby celková chyba MSE nebo MAE byla minimální.

MSE - průměrná kvadratická chyba

MAE - průměrná absolutní chyba

Machine Learning College

Lineární regrese - více vstupních atributů

vstup

plocha	patro	počet místností
42	7	2
105	3	3
67	1	2
224	3	4

výstup

cena (mil. Kč)
3,2
6,8
4,1
13,9

f(plocha, patro, pocet_mistnosti) =
w_1.plocha + w_2.patro + w_3.pocet_mistnosti + w_0

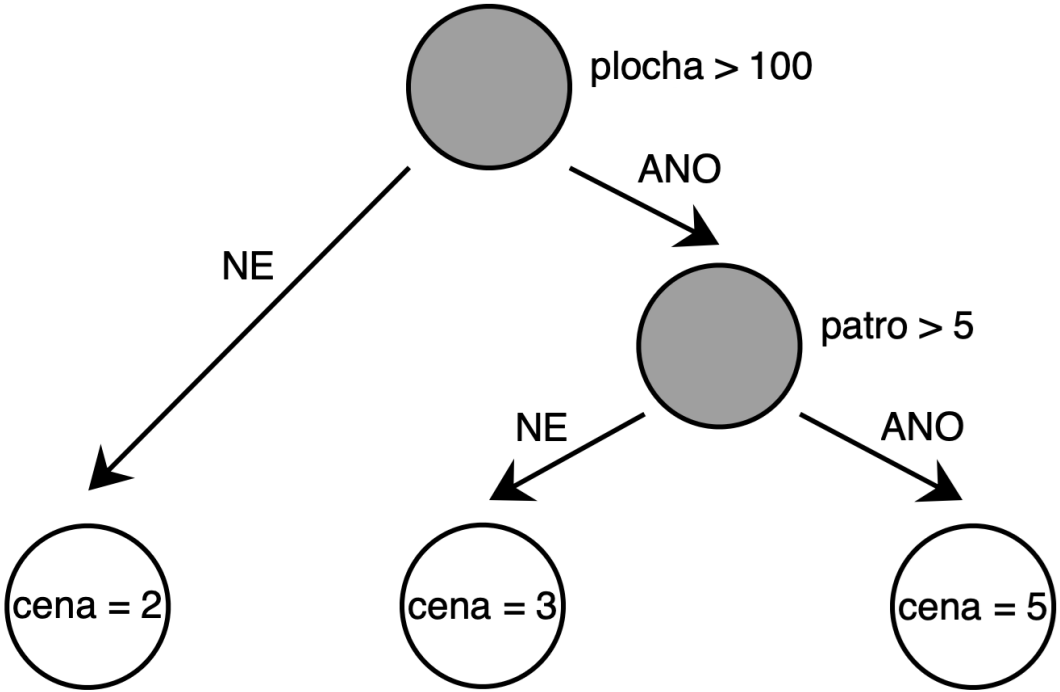
Machine Learning College

Lineární regrese - příklad

Příklad lineární regrese v knihovně ScikitLearn

Machine Learning College

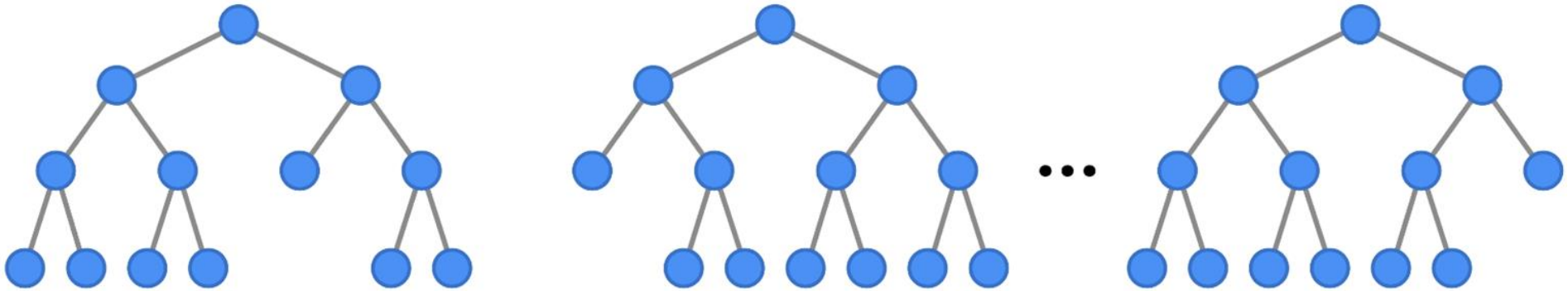
Rozhodovací strom pro regresi



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

Machine Learning College

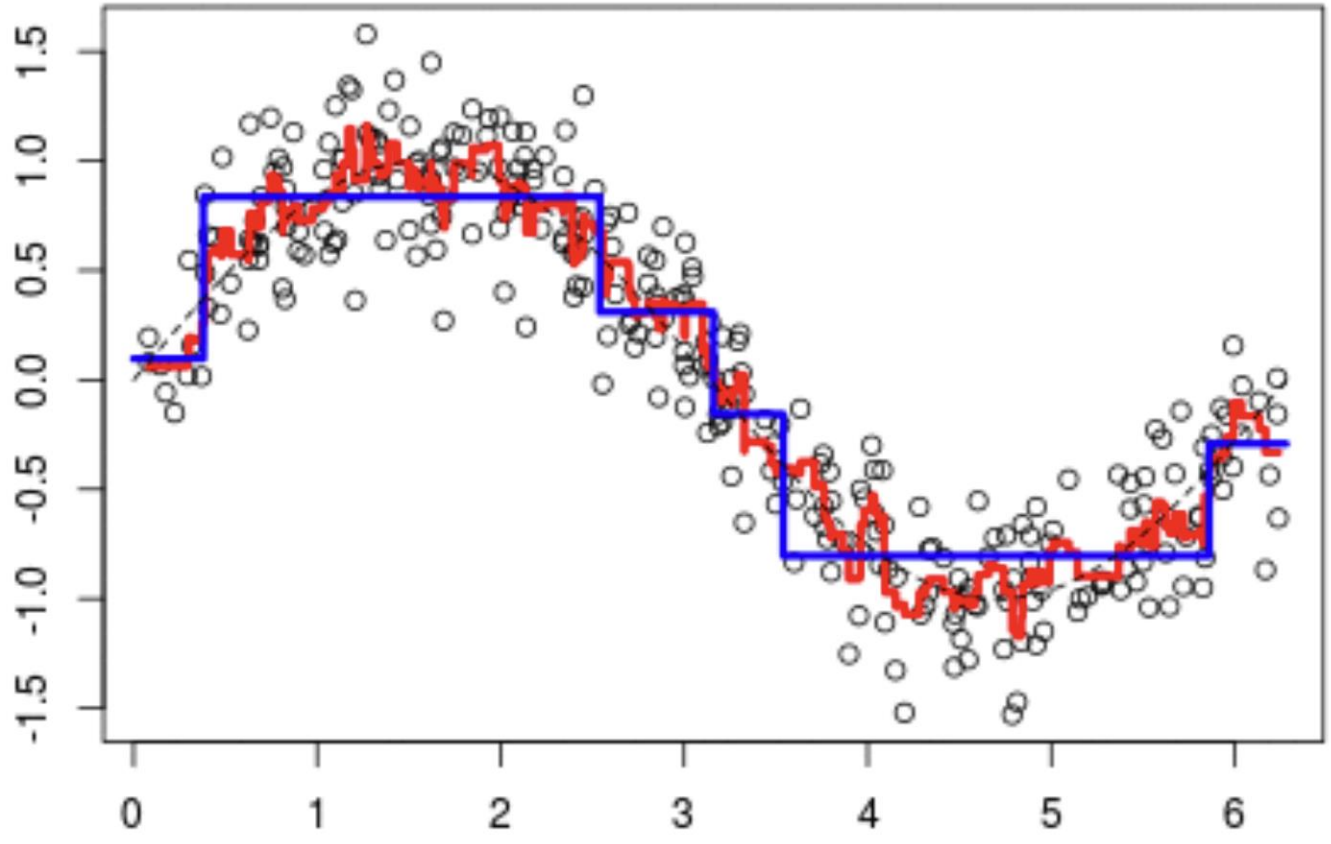
Více rozhodovacích stromů - les



- Náhodný les kombinuje více rozhodovacích stromů průměrováním jejich predikcí.
- Jednou z nejčastějších implementací je Random forest (náhodný les).
- U random forest je každý strom vytvořen z náhodně vybrané podmnožiny trénovacích dat, proto je každý strom jiný.

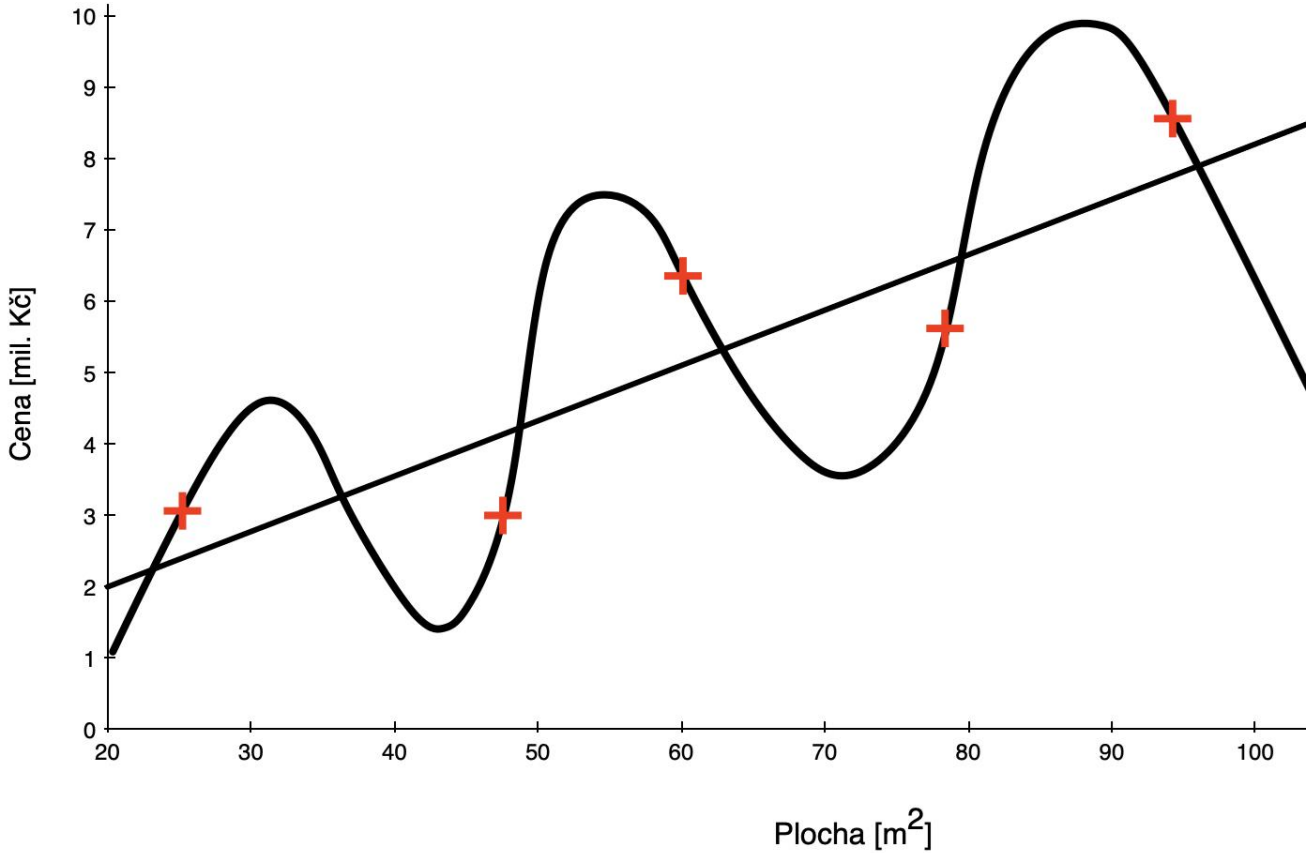
Machine Learning College

Čím více stromů, tím hladší funkce



Machine Learning College

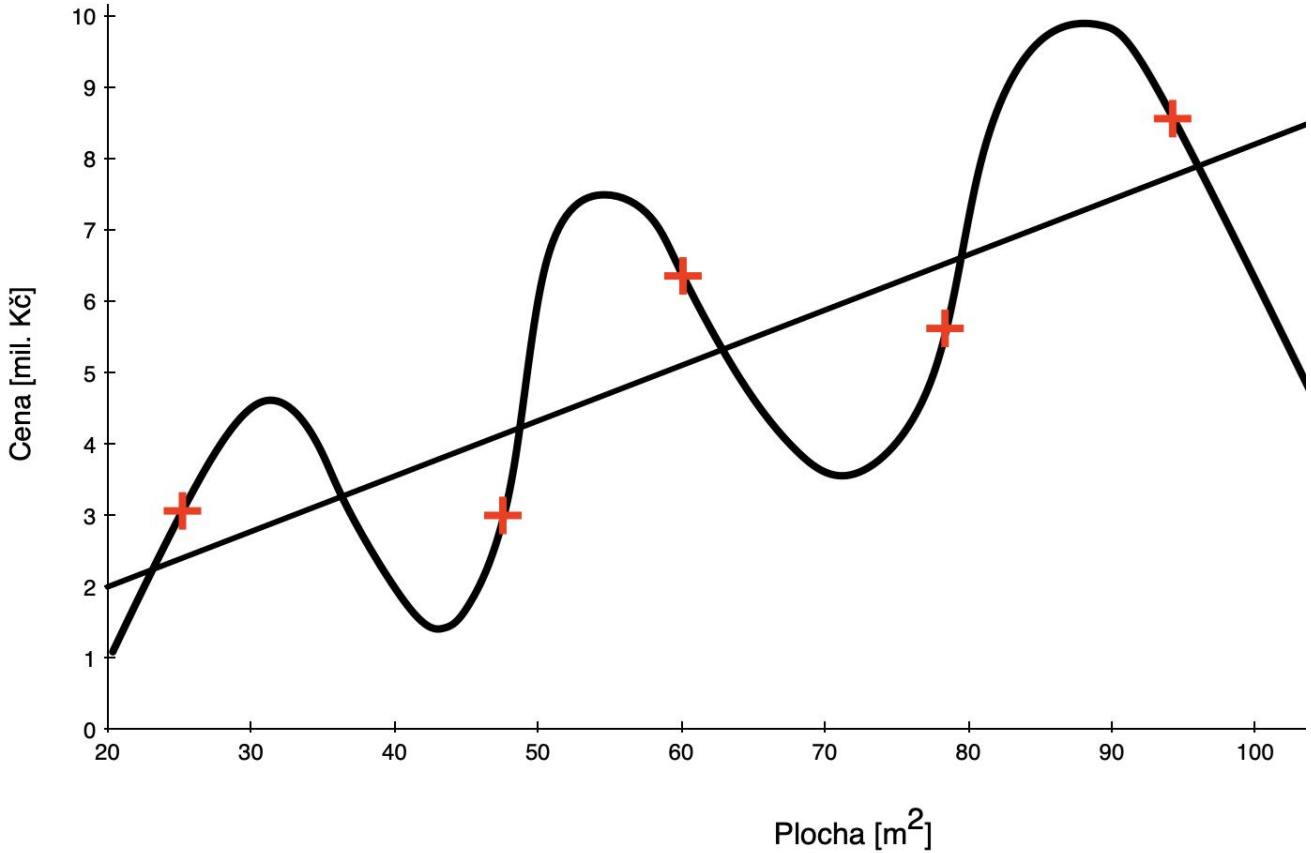
Problém přetrénování (overfitting)



- Přetrénování - model si “zapamatoval” trénovací data, ale nemá schopnost zobecnění.
- Chyba testovacího data setu je výrazně vyšší než chyba trénovacího data setu.

Machine Learning College

Základní techniky pro zabránění přetrénování



- Použití větších trénovacích dat.
- Snížení složitosti modelu (menší hloubka stromů u rozhodovacích stromů).

Machine Learning College

Volba vhodného modelu

- Různých modelů je celá řada (nejen lineární regrese a random forest).
- Je třeba zohlednit složitost problému, velikost dostupných dat, apod.
- Vždy je dobré vyzkoušet více modelů a začínat od jednodušších.

Machine Learning College