

Lineární regrese

Reprezentace
modelu

Strojové učení

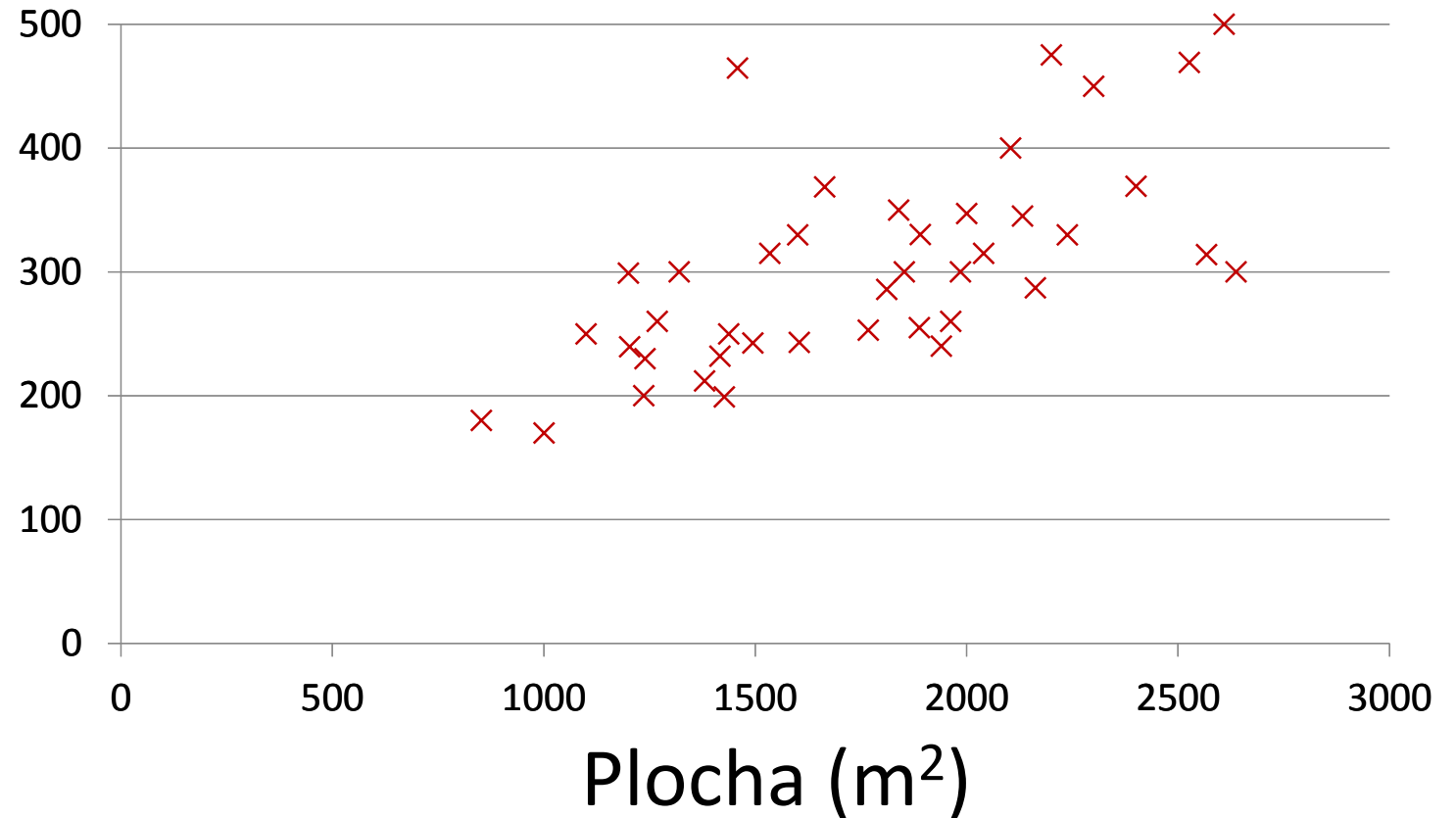
- Schopnost počítače učit se bez explicitního naprogramování
- Supervizované učení
 - Algoritmus se učí na známých datech
 - Regrese
- Nesupervizované učení
 - Algoritmus se učí na neznámých datech
 - Klasifikace
- Trénovací sada
 - Soubor dat/bodů/objektů, na kterých se algoritmus učí

Obecné vlastnosti lineární regrese

- Statistická metoda
 - Kvantifikuje závislost mezi dvěma spojitými proměnnými
 - Závislá – proměnná, kterou predikujeme
 - Nezávislá – prediktivní proměnná
- Model lineární regrese je reprezentován přímkou
 - Prochází mezi jednotlivými body
 - Součet druhých mocnin odchylek od každého bodu je minimální
- Nelineární vztah mezi proměnnými
 - Neprokládáme body přímkou, ale vhodnou křivkou
- Nezávislých proměnných může být mnoho
 - V kurzu řešíme pouze případ s jednou nezávislou proměnnou

Příklad cen domů

Cena
(v 1000 \$)



Regresní úloha

Predikujeme číselnou hodnotu závislé proměnné.

Klasifikační úloha

Predikujeme třídu, do které bod patří.

Sada bodů s cenami domů	Plocha v m² (x)	Cena v 1000 \$ (y)
	2104	460
	1416	232
	1534	315
	852	178

Značení:

m = Počet bodů v trénovací sadě

x = vstupní proměnná / atribut

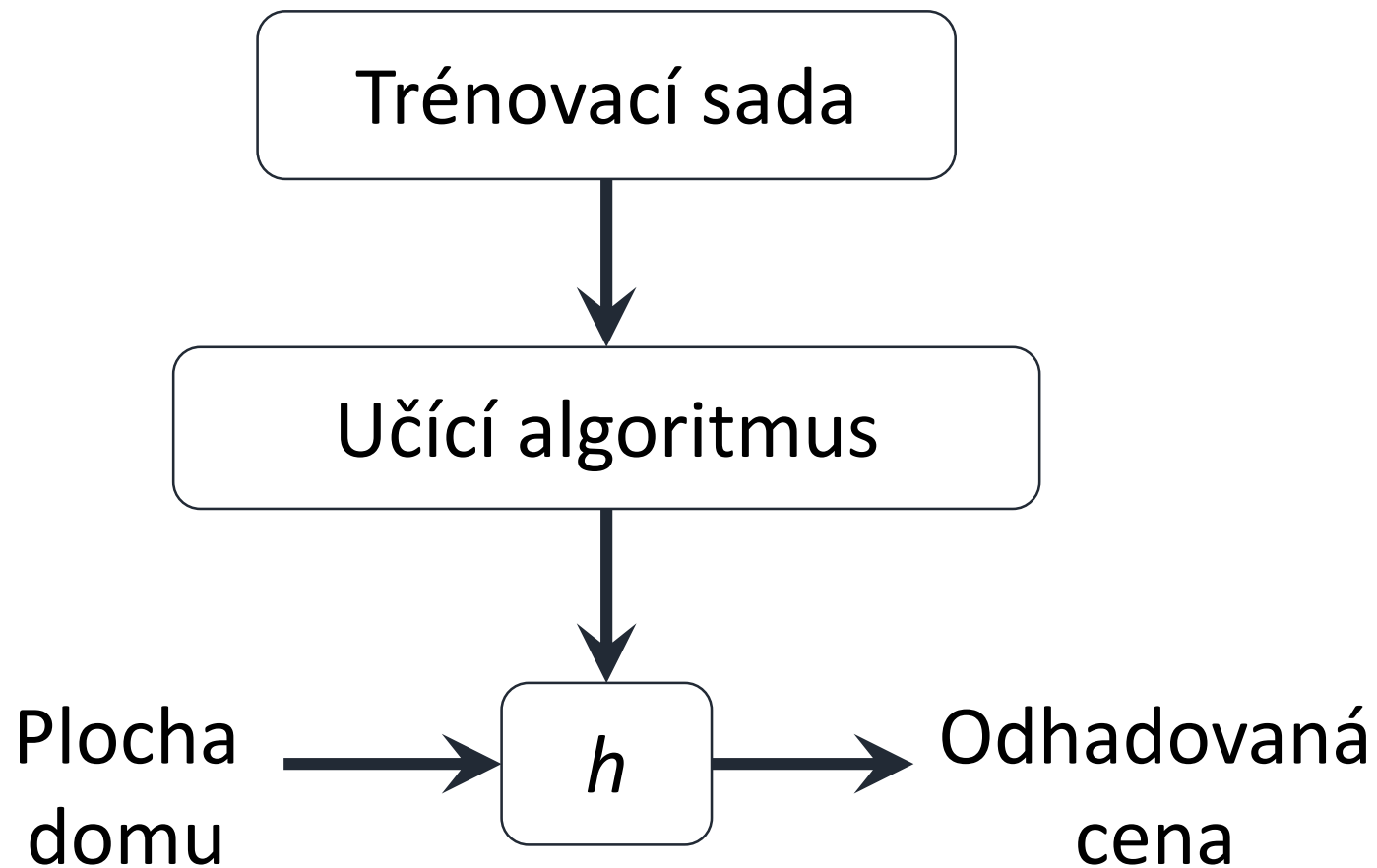
y = výstupní proměnná / cílová proměnná

(x, y) - jeden bod z množiny obecně

$(x^{(i)}, y^{(i)})$ - konkrétní i -tý bod z množiny

$$x^{(1)} = 2104$$

$$y^{(4)} = 178$$



Jak vypadá h ?

$$h_{\theta}(x) = \theta_0 + \theta_1 x$$

Jak nalézt parametry θ ?

- Učící algoritmus je v našem případě lineární regrese