

Regrese (Regression)

Regrese je statistická a strojová metoda, jejímž cílem je předpovědět **spojitou numerickou hodnotu** (číslo) na základě vstupních dat. Na rozdíl od klasifikace, která zařazuje data do kategorií (např. „pes“ vs. „kočka“), regrese odpovídá na otázku „Kolik?“.

1. Základní princip

V regresi hledáme matematickou funkci $f(x)$, která nejlépe popisuje vztah mezi nezávisle proměnnými (vstupy, rysy) a závisle proměnnou (výstupem, cílem).

$$y = f(x) + \epsilon$$

Kde:

- y je hodnota, kterou chceme předpovědět (např. cena nemovitosti).
- x jsou vstupní data (např. rozloha bytu, lokalita, stáří).
- ϵ je náhodná chyba (šum).

2. Typy regresních modelů

A. Lineární regrese (Linear Regression)

Nejjednodušší forma regrese, která předpokládá lineární vztah mezi vstupy a výstupem. Výsledkem je přímka (v 2D) nebo nadrovina (ve více dimenzích).

- **Jednoduchá:** Predikce na základě jednoho faktoru (např. plocha → cena).
- **Vícenásobná:** Predikce na základě více faktorů (plocha + lokalita + patro → cena).

B. Polynomiální regrese

Používá se, pokud data nevykazují lineární trend, ale spíše křivku. Do rovnice se přidávají mocniny vstupních proměnných (x^2 , x^3 , \dots).

C. Logistická regrese (Pozor na záměnu!)

Navzdory názvu se v ML častěji používá pro **klasifikaci**. Používá sigmoidní funkci k určení pravděpodobnosti, že daný jev nastane (výsledek je v intervalu 0 až 1).

D. Regresní stromy a lesy (Random Forest Regression)

Využívají rozhodovací stromy k rozdělení dat do skupin a následně průměrují hodnoty v těchto

skupinách. Jsou velmi odolné vůči odlehlým hodnotám (outliers).

3. Metriky hodnocení (Jak poznat dobrou regresi?)

U regrese nemůžeme měřit „přesnost“ (accuracy) jako u klasifikace, protože trefit se přesně na desetinné místo je téměř nemožné. Používáme proto metriky založené na chybě:

- **MAE (Mean Absolute Error):** Průměrná absolutní chyba. Říká nám, o kolik se v průměru model plete (např. „model se plete o 500 Kč“).
- **MSE (Mean Squared Error):** Průměrná čtvercová chyba. Více penalizuje velké chyby.
- **RMSE (Root Mean Squared Error):** Odmocnina z MSE. Má stejné jednotky jako cílová proměnná.
- **R-squared (R^2):** Koeficient determinace. Říká, jak velké procento rozptylu dat model vysvětluje (0 až 1).

4. Praktické příklady využití

- **Finance:** Předpovídání vývoje cen akcií nebo měnových kurzů.
- **Reality:** Odhad tržní ceny nemovitosti na základě jejích parametrů.
- **Meteorologie:** Předpověď teploty na následující dny.
- **Průmysl:** Prediktivní údržba (odhad, za kolik hodin se stroj porouchá).

5. Úskalí regrese

- **Overfitting (Přeučení):** Model se naučil trénovací data „nazpaměť“ i se šumem a selhává na nových datech.
- **Underfitting (Nedostatečné naučení):** Model je příliš jednoduchý (např. snaží se popsat složitou křivku přímkou).
- **Outliers (Odehlé hodnoty):** Extrémní hodnoty v datech mohou lineární regresi drasticky zkreslit.

Související články:

- [Učení s učitelem](#)
- [Základy statistiky](#)
- [Overfitting a Underfitting](#)

Tagy: ml algoritmy statistika regrese math

From:

<https://serviceit.cz/> - **IT ENCYKLOPEDIE**

Permanent link:

<https://serviceit.cz/doku.php?id=regrese>

Last update: **2026/01/02 12:41**

