**Regresszió**

A regresszió feladata az, hogy a függő változót (célváltozó, y) egy vagy több független változó (input változók, X) alapján előre jelezze. A cél itt a folytonos értékek előrejelzése, például házárak, hőmérséklet, vagy a részvényárfolyamok. A regresszió leggyakoribb típusa a lineáris regresszió, de számos más típus is létezik, mint például a polinomiális regresszió, a logisztikus regresszió (bár ez gyakran a klasszifikációval azonosítják) és a Ridge regresszió.

**Klasszifikáció**

A klasszifikáció célja, hogy az input adatok alapján egy adott kategóriába sorolja a megfigyeléseket. Itt a célváltozó diszkrét értékekkel bír, például az email spam vagy nem spam kategóriába sorolása, vagy a betegek diagnosztizálása egy adott betegségre vagy annak hiányára. A klasszifikáció algoritmusai közé tartozik a logisztikus regresszió, a döntési fák, a támogatott vektor gépek (SVM) és a neurális hálózatok.

**Különbségek**

* **Kimeneti változó típusa:** Regresszió esetén folytonos, klasszifikáció esetén diszkrét.
* **Feladat típusa:** Regresszió előrejelzést készít, klasszifikáció osztályoz.
* **Értékelési metrikák:** Regresszióban gyakori metrikák az MSE, RMSE, MAE, míg klasszifikációban a pontosság, F1-score, ROC-AUC.

**Egy- és többváltozós lineáris regresszió**

**Hipotézisfüggvény**

**Egyszerű (egyváltozós) lineáris regresszió**  
A hipotézisfüggvény egy lineáris összefüggést ír le a független változó (bemenet) és a függő változó (kimenet) között. Az egyszerű lineáris regresszió egyetlen független változót használ a célváltozó előrejelzéséhez. A hipotézisfüggvény a következőképpen néz ki:  
**hθ(x)=θ0+θ1x**

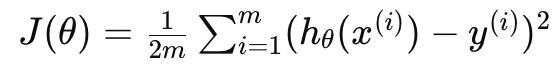
* **hθ(x):** Az előrejelzett érték
* **θ0:** Intercept - Metszéspont
* **θ1:** Slope - Meredekség
* **x:** Bemeneti változó

**Többváltozós (multivariable) lineáris regresszió**  
A többváltozós lineáris regresszió esetén több bemeneti változó is szerepel a modellben, ezáltal a hipotézisfüggvény kiterjed a több dimenziós esetre:

[Többváltozós lineáris regresszió](https://github.com/Miki0195/4.Felev/blob/main/Neuronh%C3%A1l%C3%B3k/vizsga/tobbvaltozoslinearisregresszio.png)

* **x0, x1, ... ,xn:** Bemeneti változók
* **θ0, θ1, ... ,θn:** Modell paraméterei

**Költségfüggvény**

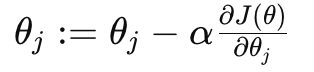
A költségfüggvény a predikciók és a valóságos értékek közötti különbség négyzetesének átlagát vagy más módon számított hibát méri. A négyzetes hiba (Mean Squared Error, MSE) gyakran alkalmazott költségfüggvény az egyváltozós lineáris regresszióban.[](https://github.com/Miki0195/4.Felev/blob/main/Neuronh%C3%A1l%C3%B3k/vizsga/Negyzeteshibaosszeg.png)

* **m:** Az adatok száma
* **hθ(xi):** Az **i**-edik adatpont előrejelzése
* **yi:** Az **i**-edik adatpont valódi értéke

**Megoldás gradiens módszerrel**

A gradiens módszer (Gradient Descent) egy iteratív optimalizációs algoritmus, amely célja a költségfüggvény minimalizálása. Az algoritmus a következőképpen működik:

* **Kezdeti paraméterek választása:** Véletlenszerűen vagy nullával.
* **Gradiens számítása:** Számítsuk ki a költségfüggvény gradientjét a paraméterek szerint.
* **Paraméterek frissítése:** Frissítsük a paramétereket a tanulási ráta (α) és a gradiens alapján.

Az egyes paraméterek frissítése: [](https://github.com/Miki0195/4.Felev/blob/main/Neuronh%C3%A1l%C3%B3k/vizsga/gradiensmodszer.png)

**Alkalmazások**

Alkalmazásukban nagy különbés nincs, többváltozós lineáris regressziót akkor használjuk inkább, amikor több bemeneti változó határozza meg a kimeneti változót, például házárak becslése során, amikor több tényező is befolyásolhatja az árat.

A lineáris regresszió széles körben alkalmazott számos területen, például:

* **Gazdasági előrejelzések:** Például részvényárfolyamok vagy GDP előrejelzése.
* **Egészségügy:** Például a vérnyomás előrejelzése a kor, testsúly és egyéb jellemzők alapján.
* **Marketing:** Hirdetési költségek és az eladások közötti kapcsolat vizsgálata.
* **Tudományos kutatások:** Különböző kísérleti adatok analízise.