**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**

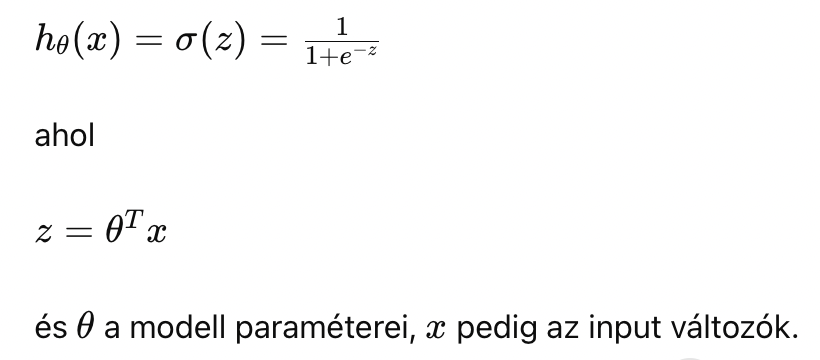
* **Kimeneti változó típusa:** Regresszió esetén folytonos, klasszifikáció esetén diszkrét.
* **Feladat típusa:** Regresszió előrejelzést készít, klasszifikáció osztályoz.
* **Értékelési metrikák:** Regresszióban gyakori metrikák az MSE, RMSE, MAE, míg klasszifikációban a pontosság, F1-score, ROC-AUC.

**Logisztikus Regresszió**

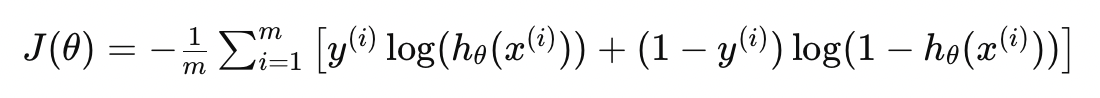
A logisztikus regresszió egy statisztikai modell, amelyet bináris és többosztályos klasszifikációs feladatokra használnak. Az alapja a logisztikus függvény, amely az input változókat valószínűségekké alakítja.

**Egyszerű Logisztikus Regresszió**

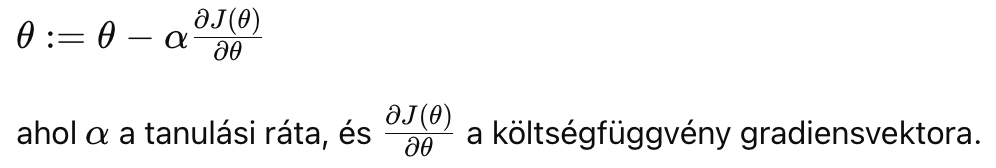
**Hipotézisfüggvény**

A logisztikus regresszió hipotézisfüggvénye a logisztikus (sigmoid) függvény, amely a bemeneti adatokat egy 0 és 1 közötti valószínűségbe transzformálja. [](https://github.com/Miki0195/4.Felev/blob/main/Neuronh%C3%A1l%C3%B3k/vizsga/adom.png)

**Költségfüggvény**

A logisztikus regresszió költségfüggvénye a log-veszteség (log loss), amely mérni hivatott, mennyire jól illeszkedik a modell a tanító adatokra.[](https://github.com/Miki0195/4.Felev/blob/main/Neuronh%C3%A1l%C3%B3k/vizsga/adom3.png)

**Megoldás Gradiens Módszerrel**

A gradiens módszer segítségével a paramétereket úgy frissítjük, hogy minimalizáljuk a költségfüggvényt. A paraméterek frissítési szabálya:[](https://github.com/Miki0195/4.Felev/blob/main/Neuronh%C3%A1l%C3%B3k/vizsga/adom2.png)

**Többváltozós Logisztikus Regresszió**

**Hipotézisfüggvény**

A többváltozós logisztikus regresszió esetén több bemeneti változót használunk a kimeneti változó előrejelzéséhez. A hipotézisfüggvény ugyanaz marad, de több jellemzőt tartalmaz: [Négyzetes hibaösszeg](https://github.com/Miki0195/4.Felev/blob/main/Neuronh%C3%A1l%C3%B3k/vizsga/adom4.png)

A **költésgfüggvény** és a **megoldás gradiens módszerrel** megegyezik az egyszerű/egyváltozós esettel!

**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.

**Egyváltozós Lineáris Regresszió**

* **Pénzügyi elemzés:** Egyváltozós lineáris regressziót alkalmazhatunk például a részvényárak vagy áruk jövőbeli alakulásának előrejelzésére.
* **Marketing:** A termékértékesítés összefüggéseit modellezhetjük, például a hirdetési költség és a termékek eladása közötti kapcsolatot.
* **Környezetvédelem:** Az egyváltozós lineáris regressziót használhatjuk például az időjárási minták elemzésére vagy az éghajlatváltozás hatásainak vizsgálatára.

**Többváltozós Lineáris Regresszió:**

* **Gazdasági előrejelzés:** Többváltozós lineáris regressziót használhatunk például a GDP növekedésének előrejelzésére, figyelembe véve több gazdasági mutatót, mint például a munkanélküliségi ráta, a fogyasztás stb.
* **Orvosi kutatás:** A betegségek kockázati tényezőit és az egészségügyi mutatók közötti összefüggéseket vizsgálhatjuk, hogy megértsük például az elhízás vagy a dohányzás egészségügyi következményeit.
* **Termelési tervezés:** A gyártási folyamatok hatékonyságának növelésére és az erőforrások optimalizálására használhatjuk a többváltozós lineáris regressziót, például a termelési költségek vagy a termelési idő becslésére.