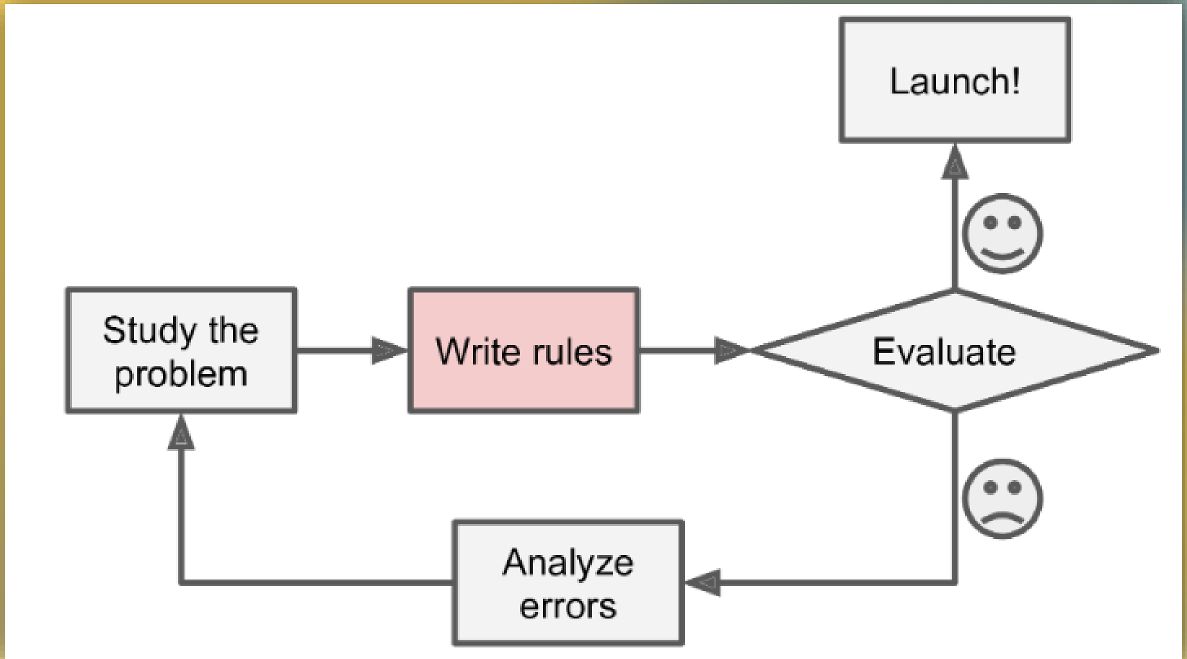
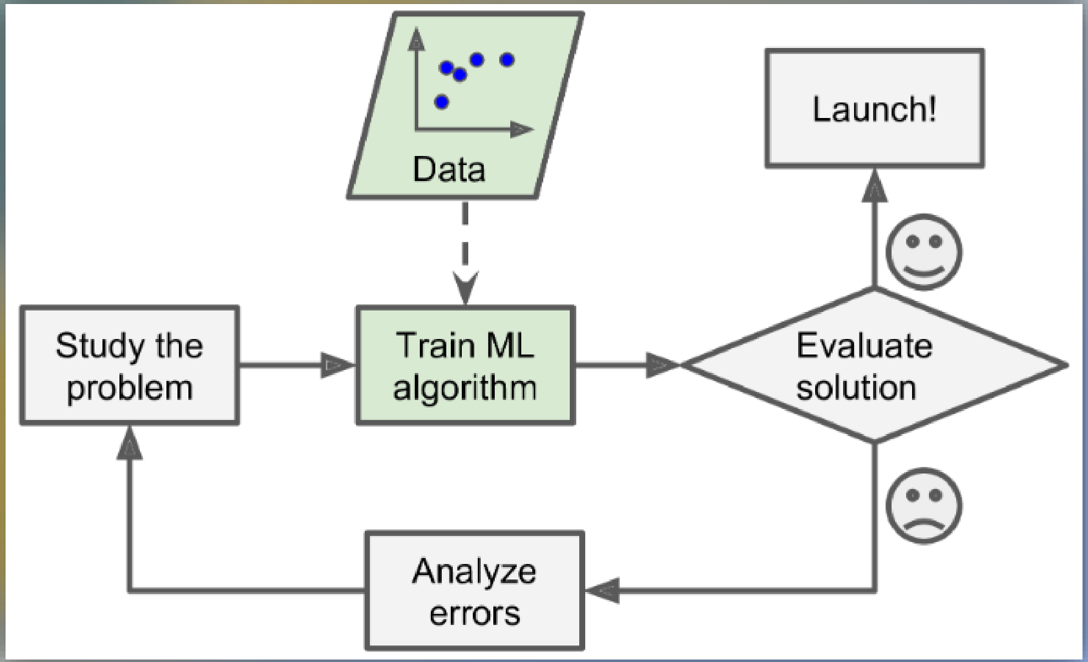
2. Előadás
Osztályozás
Logisztikus regresszió
Softmax regresszió

A determinisztikus és gépi tanulás szemlélet

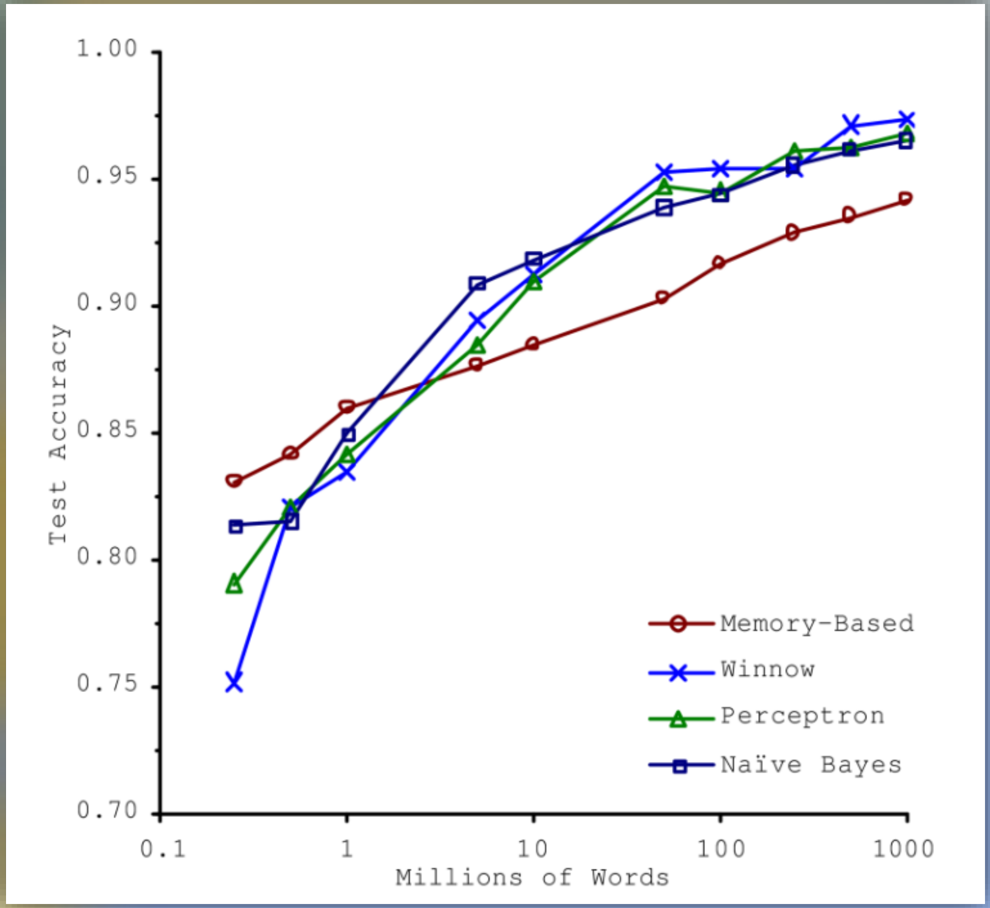




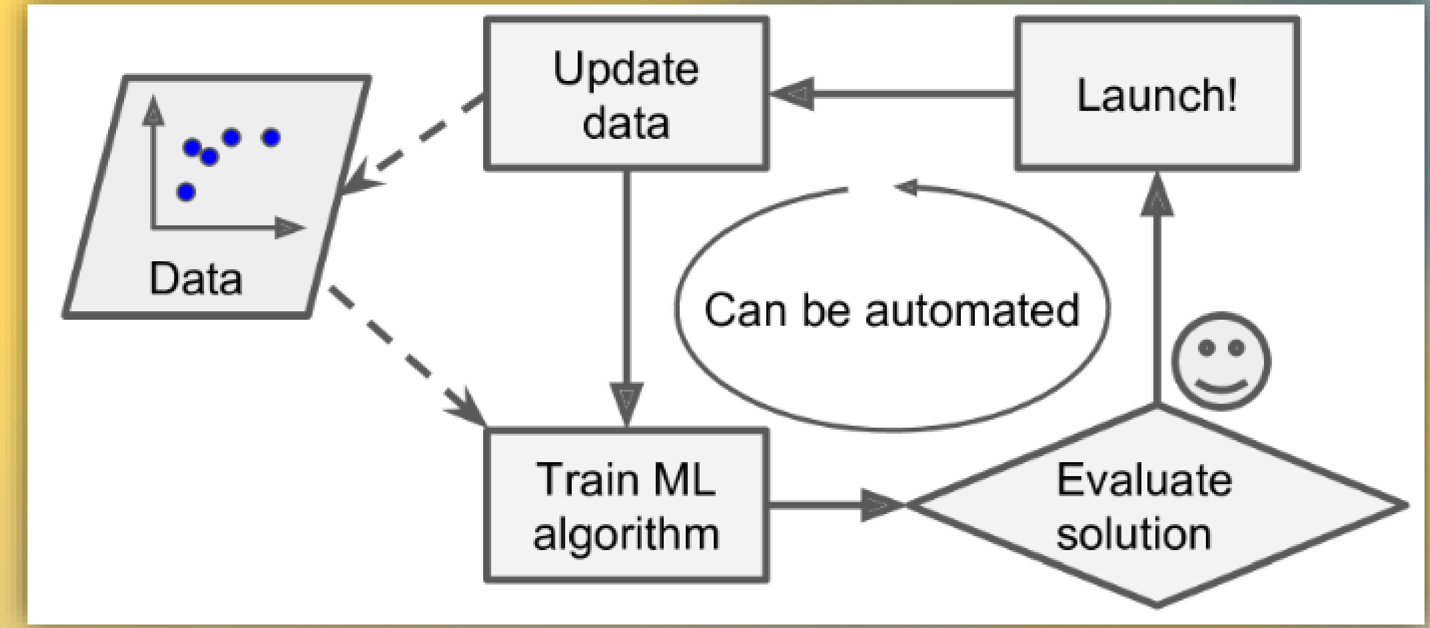
- Határozzuk meg az inputokat, outputokat, célokat és eszközöket a két irányzat szerint!

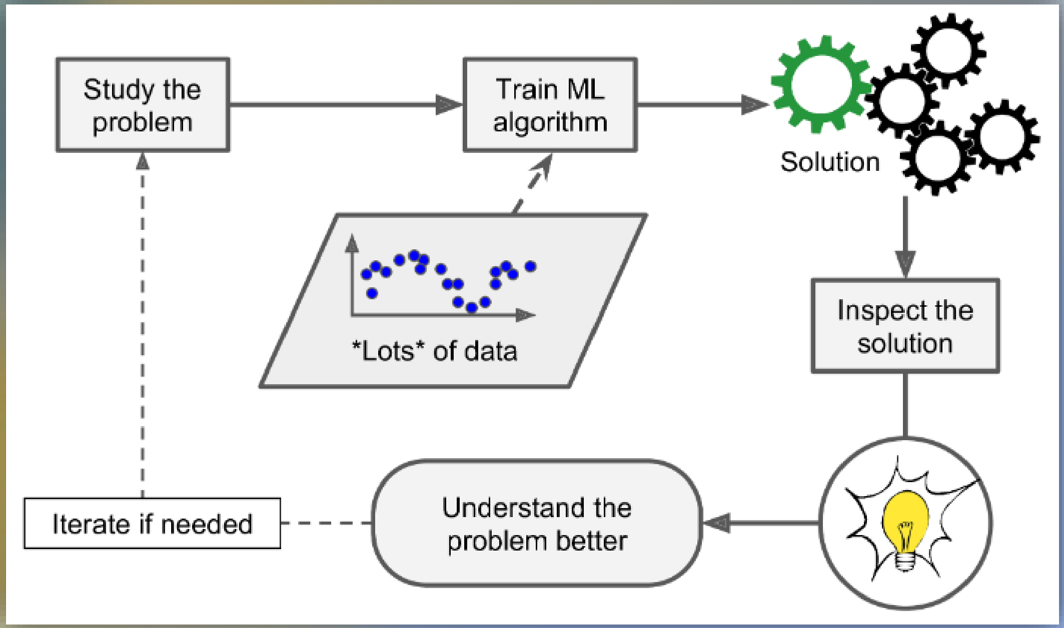
Az adatok észszerűtlen hatékonysága

- 2001-es kutatásukban Michele Blanko és Eric Brill kimutatták, hogy a különböző ML algoritmusok (az egyszerűek is) hasonlóan jól teljesítenek a természetes nyelvfelismerés területén, ha elég sok adaton tanítják a modelleket.
- Ahogy ők fogalmaztak: „*Az eredmények azt mutatják, hogy újra kell gondolnunk, mire fordítjuk a pénzünket és erőforrásainkat: algoritmusok fejlesztésére, vagy adatgyűjtésre.*”
- Miért van olyan sok béna játék a facebookon?



Tanítás automatizálása adat-alapúan





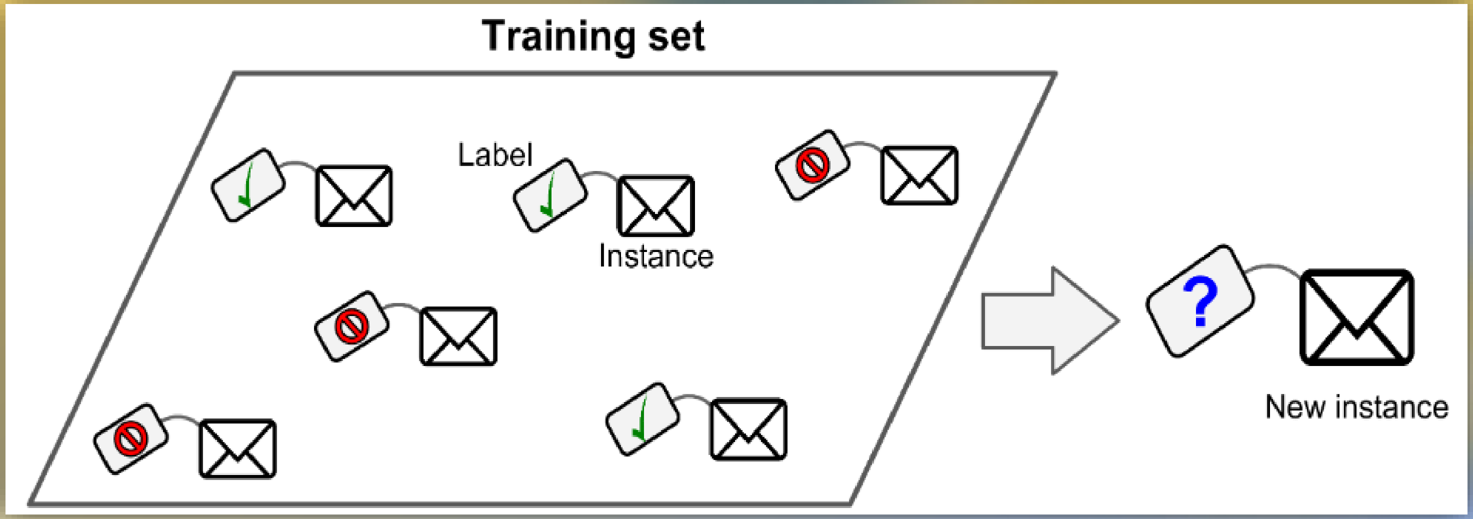
Hogyan taníthatja egymást ember és gép a gépi tanulás szemlélete szerint?

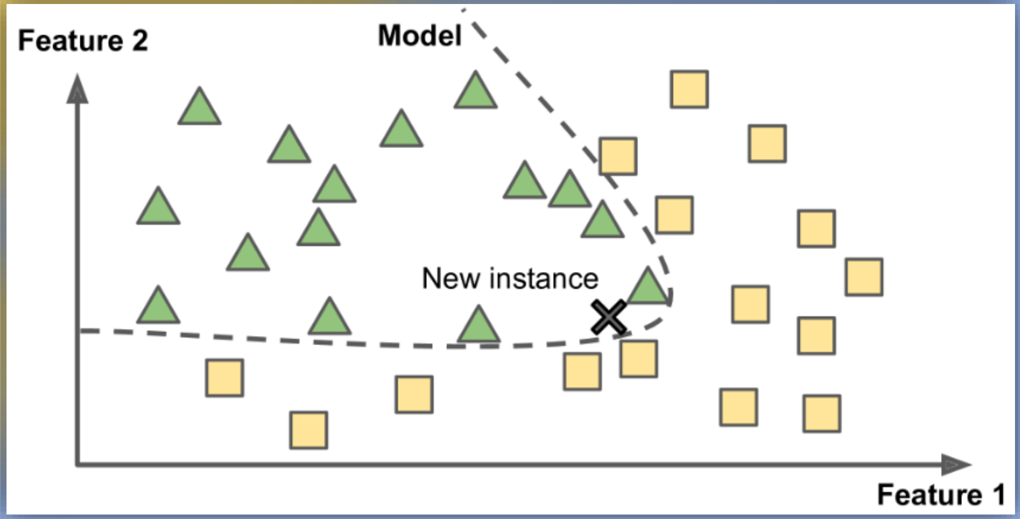
Osztályozás

- A felügyelt tanulás másik nagy irányzata a regresszió mellett.
- Osztályozásról beszélünk, amikor a predikció célváltozója kettő- vagy több, egymással átfedésben nem álló **diszjunkt kategóriára** osztható.
- Az osztályozás problémája esetén szeretnénk megismerni egy előre partícionált adathalmaz belső struktúráját. A partíciókat nevezzük kategóriának vagy osztálynak.
- Tárgya egy **diszkrét** változó.

Modellalapú tanítás

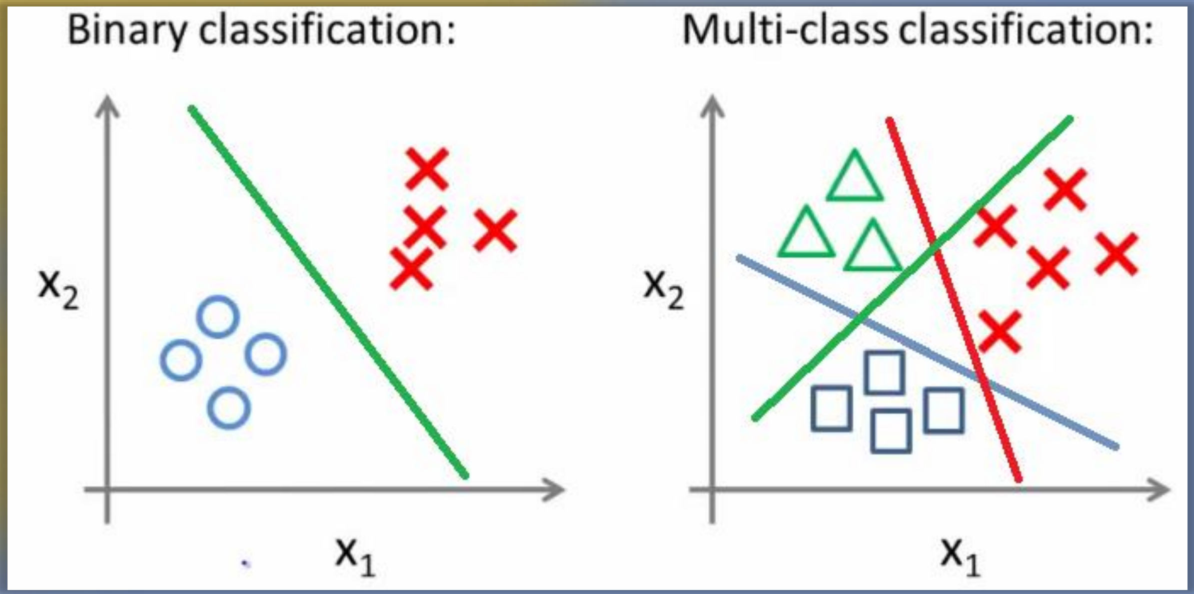
- Az osztályozó modellt arra használjuk fel, hogy eddig még nem látott mintaegyedek címkéit (osztályait) megjósoljuk vele.
- A modell inputja az osztályozási probléma esetén egy olyan adathalmaz, amely már fel van osztva partíciókra. Ez a **tanító adat/training data**.
- A modell teljesítményét egy külön erre a célra elkülönített adathalmazzal mérjük: ez a **teszt adat/testing data**.





Az osztályozás fajtái

- **Bináris**: a modell két lehetséges osztály közül valamelyikbe sorolja be az egyedeket. Minden egyedhez legalább és legfeljebb 1 osztály tartozik.
- **Multiclass**: több, mint két lehetséges kategória létezik, amibe az egyedek besorolhatók, ezek közül az egyikbe fog sorolódni az egyed. Minden egyedhez legalább és legfeljebb 1 osztály tartozik.
- **Multilabel**: minden mintaegyedhez több bináris vagy multiclass címkekategóriából tartozik elem. Pl. [Szem: *zöld*, Haj: *hosszú*]
- **Multioutput**: A multilabel osztályozás generalizált változata. Egy egyedhez egy multilabel halmazból több elem is tartozhat. Pl: [Mail: [*spam, ad, junk*]]
- Mikor a legrosszabb egy osztályozó modell?



Példa: írott számjegyek képei alapján készítsünk prediktort, ami megadja, milyen szám.

- Regresszió vagy osztályozás?

A logisztikus regresszió

- Statisztikai modellezési eljárás annak a **valószínűségnek** a megbecslésére, hogy egy adott mintaelem **valamely osztályba tartozik-e vagy sem → bináris osztályozás**.
- Ha a becsült valószínűség magasabb, mint egy adott **küszöbérték** (leggyakrabban 50%), a modell úgy dönt, hogy beletartozik a pozitív osztályba (1-es osztály).
- Ha a becsült valószínűség alacsonyabb mint 50%, a modell úgy dönt, hogy a negatív osztályba tartozik (0-s osztály).

$$\hat{y} = \begin{cases} 0 \text{ if } \hat{p} < 0.5 \\ 1 \text{ if } \hat{p} \geq 0.5 \end{cases}$$

A logisztikus függvény

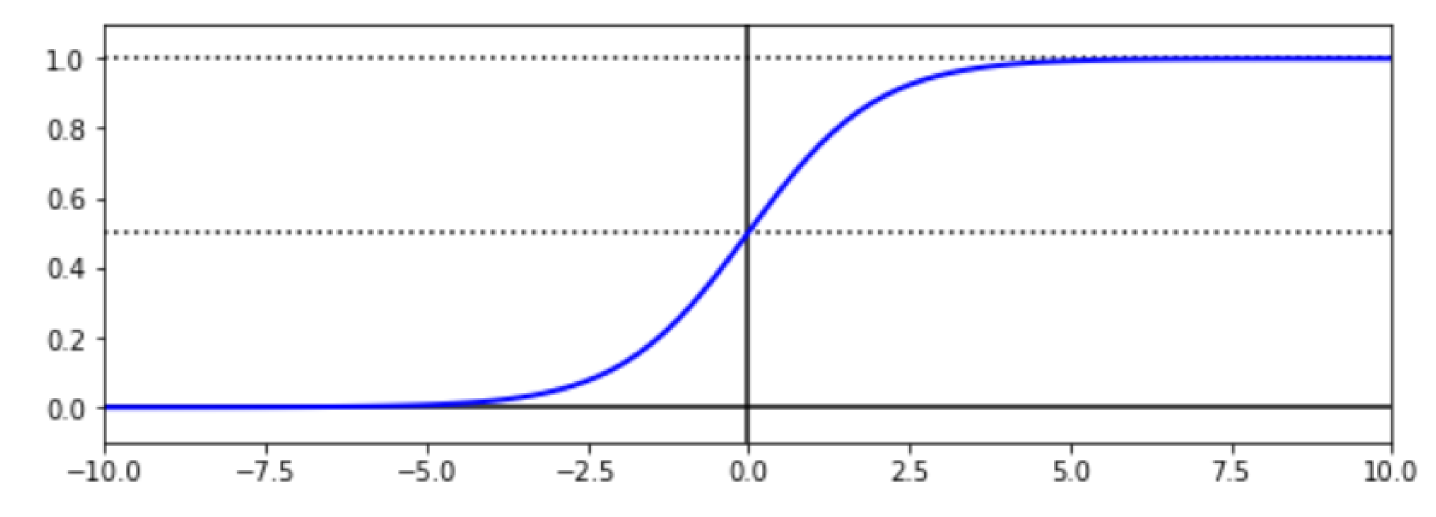
- A logisztikus modellhez felírjuk z-t a már ismert lineáris módon, hogy legyen egy index ami összevonja az adatokat.
- Majd behelyettesítjük a kapott lineáris kifejezést a logisztikus függvény jobb oldalába, hogy megkapjuk $f(z)$-t.

```
sig = 1 / (1 + np.exp(-t))
```

$$z = \underbrace{\alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k}$$

$$f(z) = \frac{1}{1 + e^{-z}}$$

$$= \frac{1}{1 + e^{-(\alpha + \sum \beta_i X_i)}}$$



De miért bukik el a lineáris regresszió diszkrét kimenetelek esetében?

- Vegyünk egy példát! Jósoljuk meg, hogy a fehér fog-e nyerni egy sakkmeccsben az alapján, hogy a két versenyző rang pontszáma között mekkora a különbség. A következő tanító tábla alapján:

Fehér és Fekete játékos rangjának különbsége	Fehér nyert?
200	0
-200	1
300	0

- Próbáljunk meg lineáris modellt felállítani az adatokra. Mit fogunk tapasztalni?

A lineáris modell bukása

- A lineáris modell a következő egyenletet eredményezi:

$$fehér\ nyert = 0.55 - 0.00214 * (Rangok\ különbsége)$$

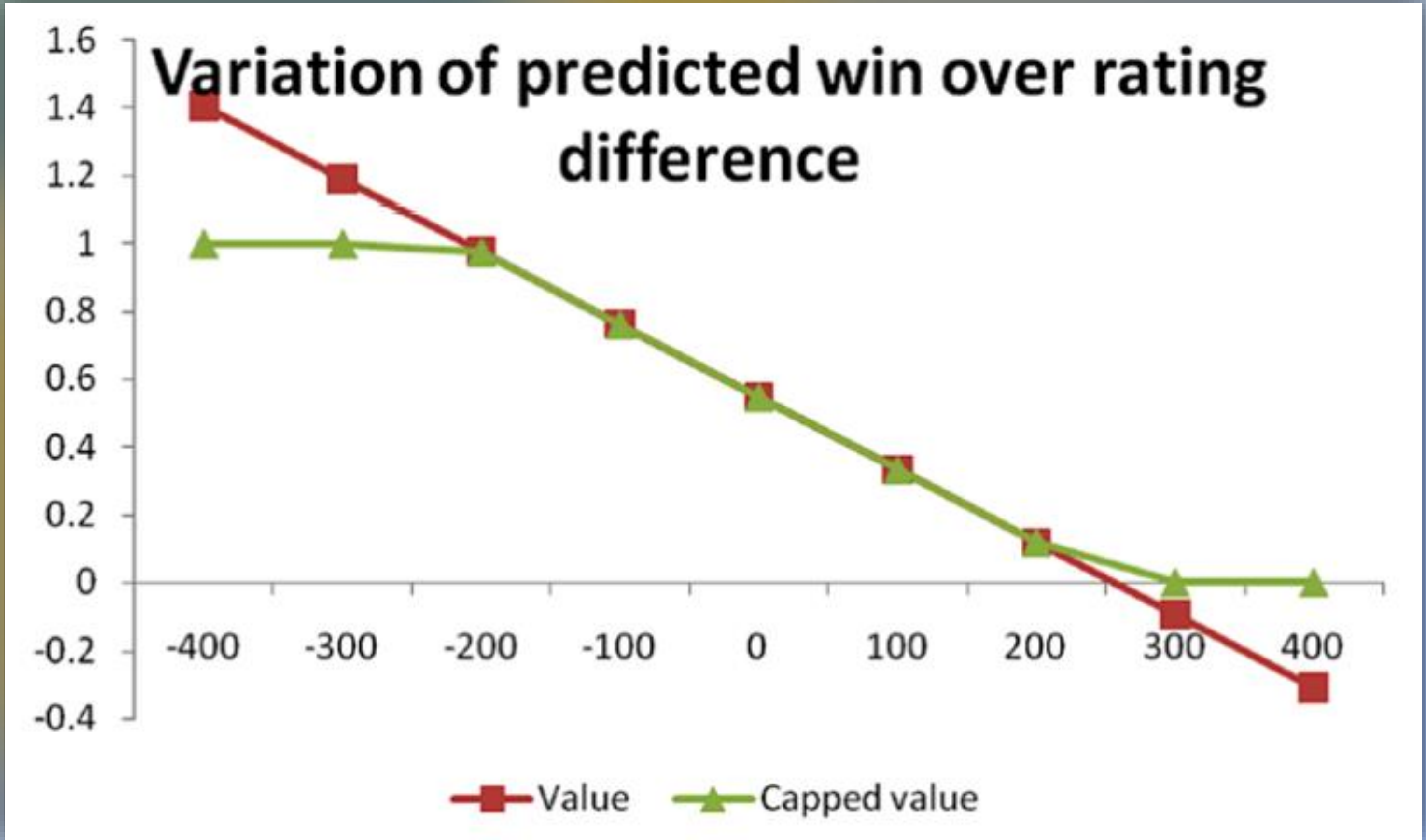
- Ezzel a modellel a következő predikciókat kapjuk:

Fehér és Fekete játékos rangjának különbsége	Fehér nyert?	Lineáris predikció
200	0	0.11
-200	1	0.97
300	0	-0.1

- Jól van ez így?
- A modellünk minden esetben racionális megoldáshoz vezet?

Hát nem!

- Ahogy látható, a 300-as különbség olyan predikcióhoz vezetett, ami kevesebb, mint 0. Hasonlóan egy -300-as különbség is valótlan értéket adna becslésül.
- Ebben az esetben a negatív értékek értelmetlenek. Negatív valószínűségek nem léteznek!
- A predikcióra használt függvény szélsőértékei 0 a ∞-ben és 1 a $-\infty$-ben.
- Az 1 feletti predikció 1 kell, hogy legyen, a 0 alatti pedig 0.



Költségfüggvény egy mintaegyedre

- A tanítás célja a θ paraméter halmaz olyan beállításának megtalálása, hogy a modell magas valószínűséggel találjon meg pozitív egyedeket: $\hat{y} = 1$
- Emellett alacsony valószínűséggel találjon meg negatív egyedeket: $\hat{y} = 0$
- Tehát minél kevesebb legyen a bizonytalalanság a modellben.
- Ez az ötlet képeződik le az egy egyedre számított költségfüggvényben:

$$c(\boldsymbol{\theta}) = \begin{cases} -\log(\hat{p}) & \text{if } y = 1 \\ -\log(1 - \hat{p}) & \text{if } y = 0 \end{cases}$$

- A $-\log(\hat{p})$ egyre magasabb, ahogy z közelít a 0-hoz, tehát a költség magas lesz, amikor a modell pozitív osztályt nagy bizonyossággal jósol negatív egyedre.
- Akkor is magas lesz, amikor nagy bizonyossággal jósol negatív osztályt pozitív egyedre.

Költségfüggvény az összes mintaegyedre

- A teljes tanító adathalmazra kiszámított költségfüggvény ($J(\theta)$) egyszerűen az egyes tanító mintaelemekre kiszámított logisztikus költségek átlaga. Egyetlen kifejezésben felírható, ez a **log loss fügvény:**

$$J(\boldsymbol{\theta}) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}log\left(\hat{p}^{(i)}\right) + \left(1 - y^{(i)}\right)log\left(1 - \hat{p}^{(i)}\right)\right]$$

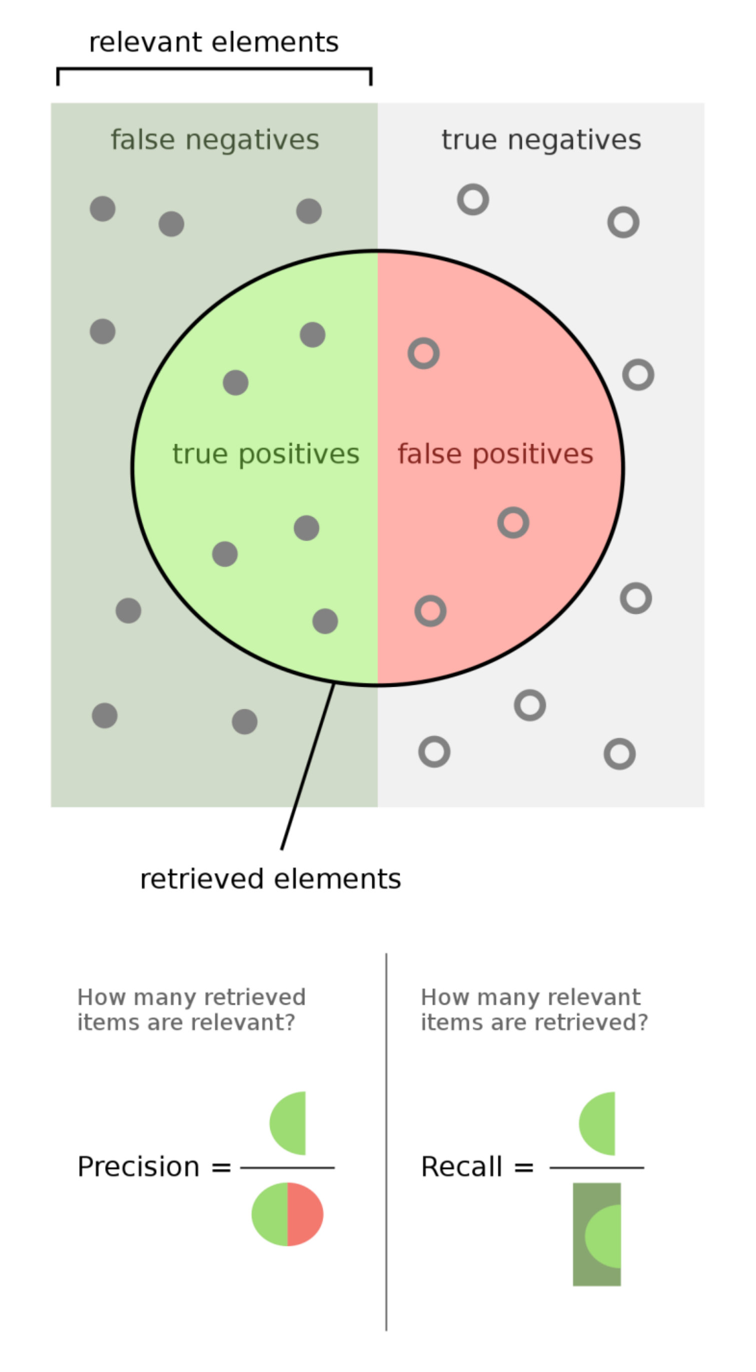
- A rossz hír: nincs olyan zárt formájú számítás, ami θ azon halmazát adja, ami minimalizálja a költségfüggvényt.
- A jó hír: a függvény **konvex**, ezért a minimum hely közelítése lehetséges ilyen eljárásokkal, pl. gradiens ereszkedéssel.
- A költségfüggvénynek deriválhatónak kell lennie. Miért?

Az osztályozás jóságának mérése

- True Positive: pozitív egyed, és annak is van osztályozva. (**TP**)
- True Negative: negatív egyed, és annak is van osztályozva. (**TN**)
- False Positive: negatív egyed, de pozitívnak van osztályozva. (**FP**)
- False Negative: pozitív egyed, de negatívnak osztályozva. (**FN**)
- Két fontos mutató: **precision** és **recall**.

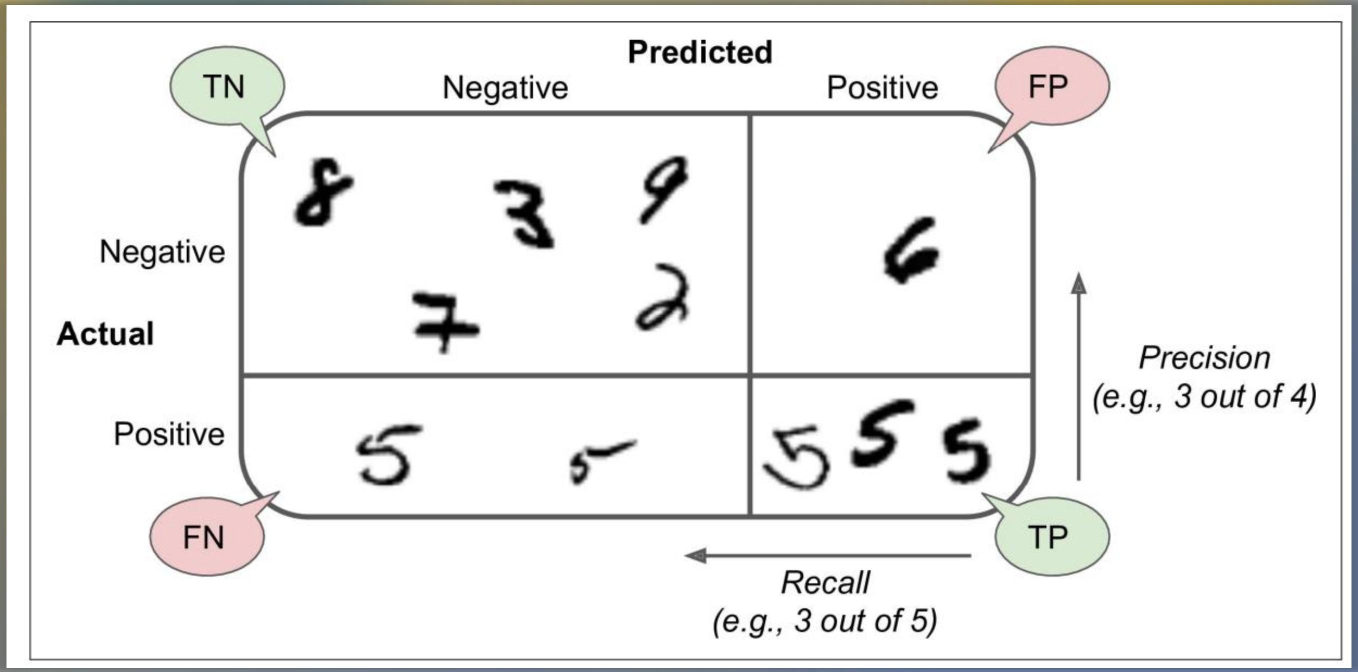
$$\text{precision} = \frac{TP}{TP + FP}$$

$$\text{recall} = \frac{TP}{TP + FN}$$



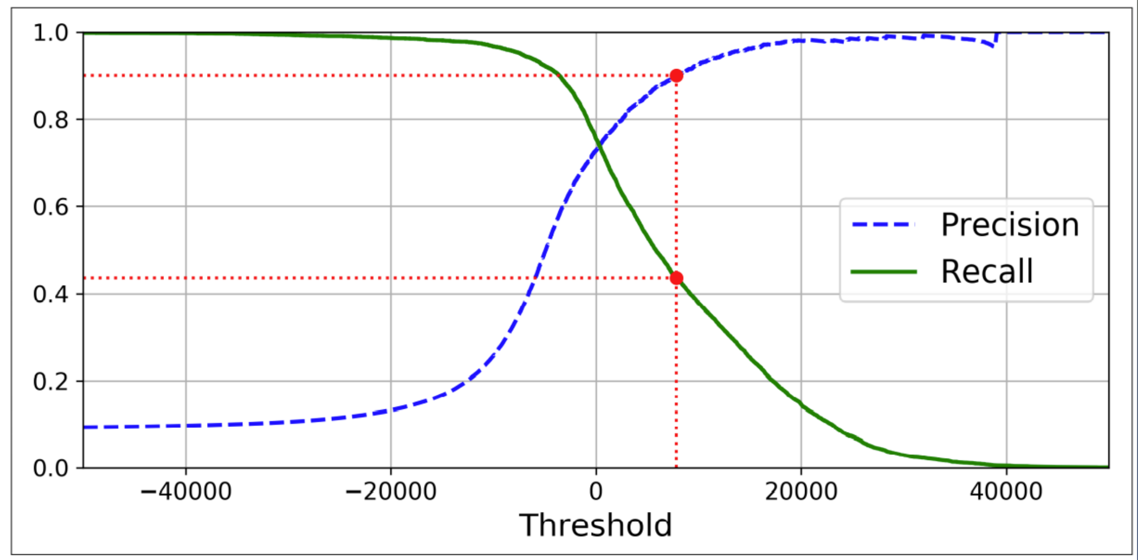
A konfúziós mátrix

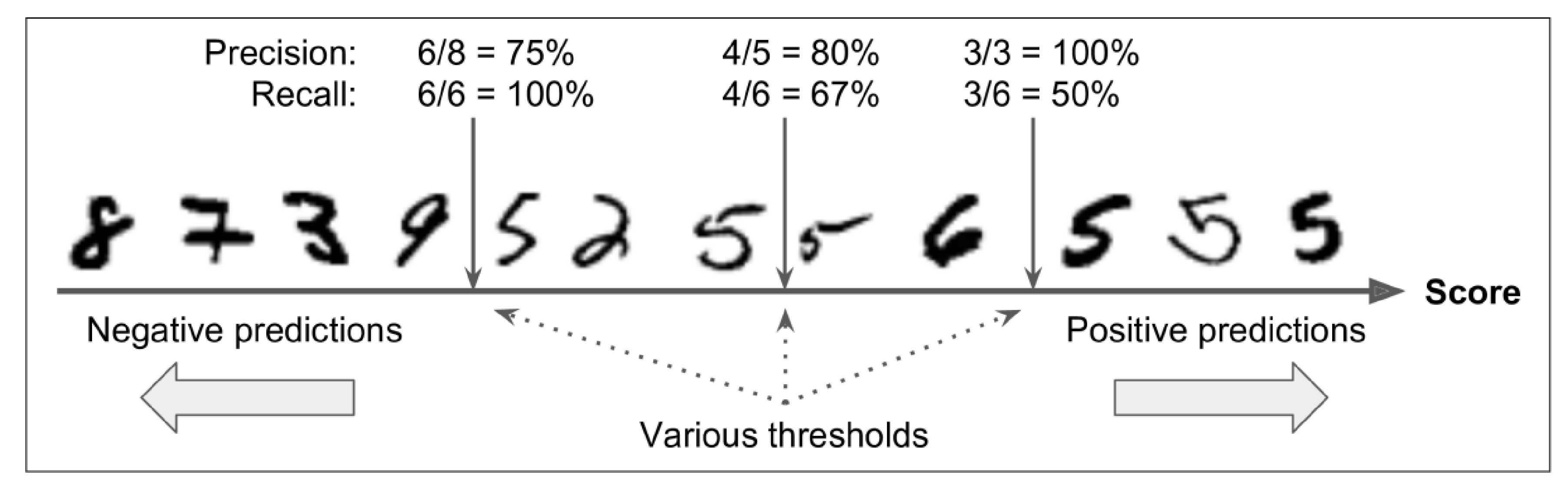
- Az osztályozás jóságának mérőeszköze, a valós és becsült osztályokat hasonlítja össze minden mintaegyedre.
- Minden sora egy valós érték, minden oszlopa bedig a becsült érték.
- Annál jobb, minél több elem található a főátlón. (Nem csak $2x2$!)



Precision / Recall Tradeoff

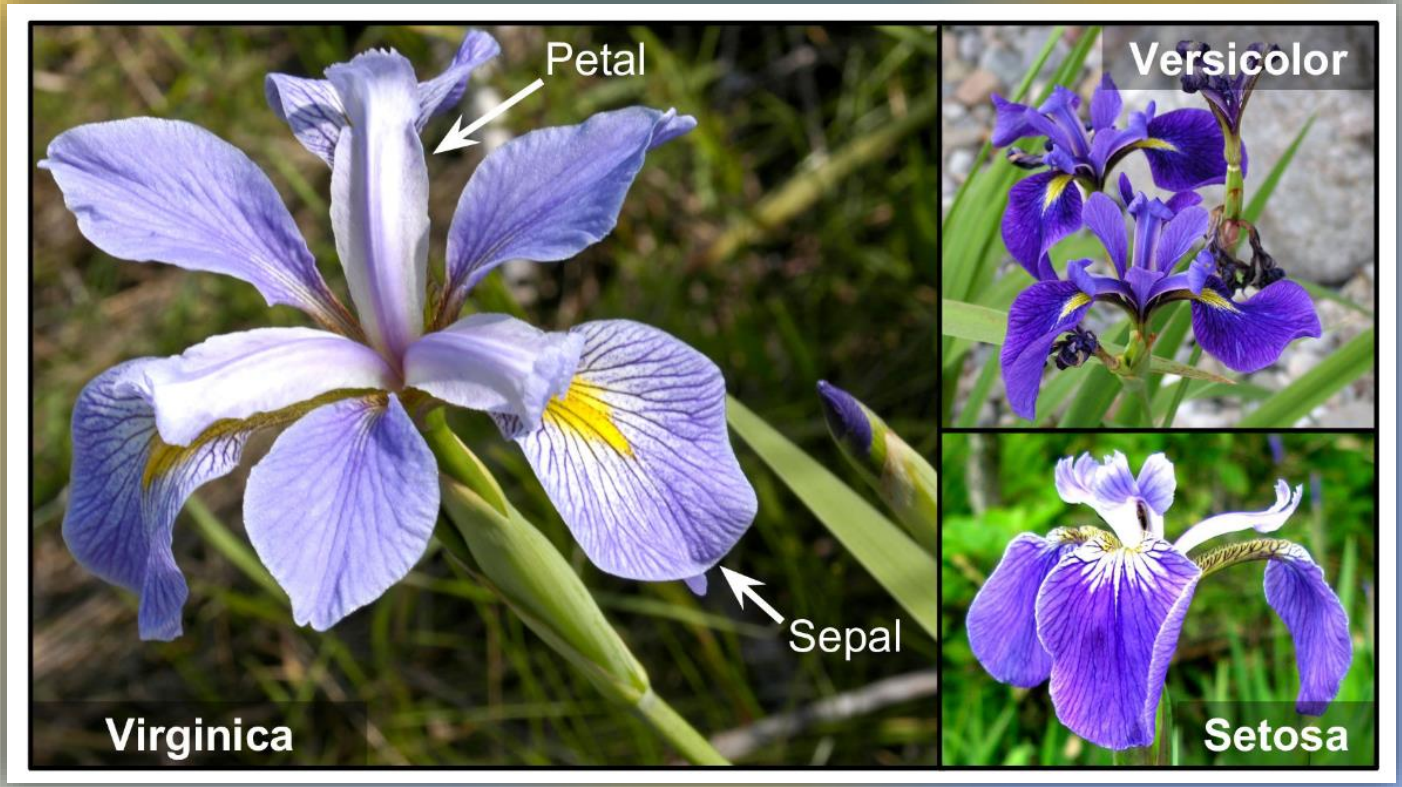
- Sajnos nem lehet mindkettőre optimalizálni osztályozás esetén.
- Vagy a precision szerint teljesít jól a modell, vagy a recall szerint.
- A döntési küszöb értékéből következik, hogy melyiket részesíti előnyben a modell.





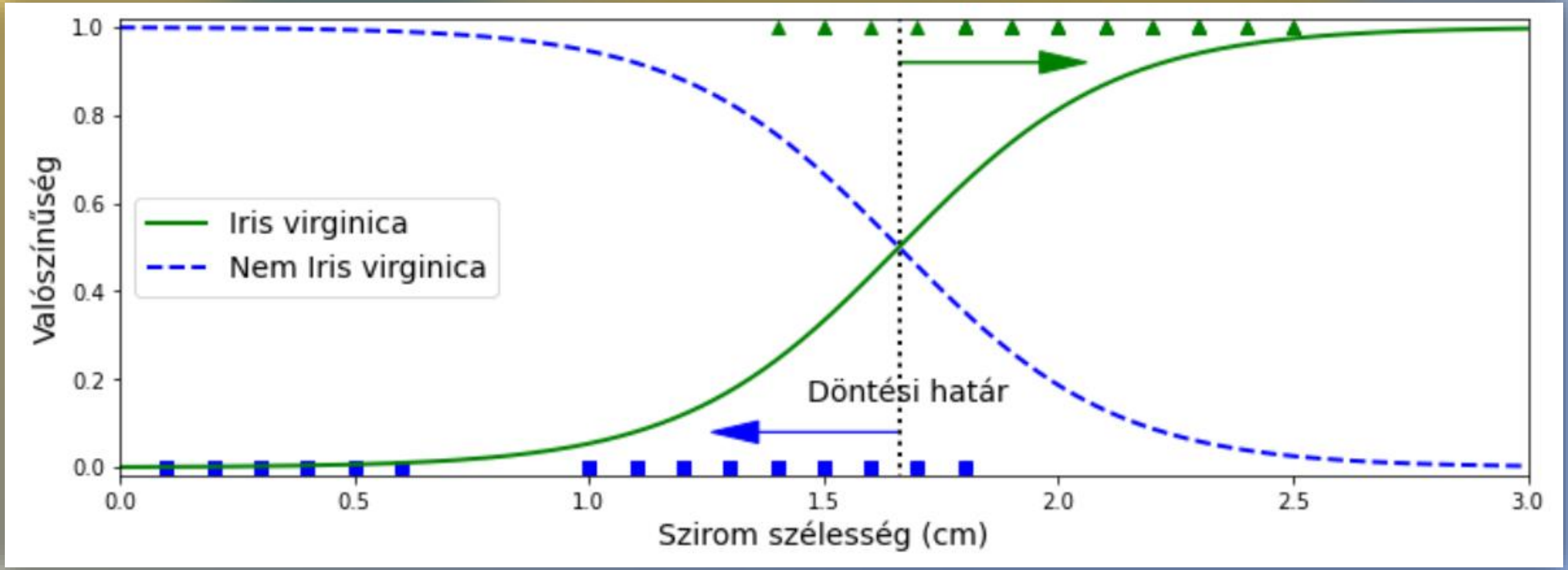
A logisztikus modell

- Használjuk fel a logisztikus regressziót arra, hogy a sziromszélesség alapján megjósoljuk, hogy egy virág az Iris Virginica osztályba tartozik-e!



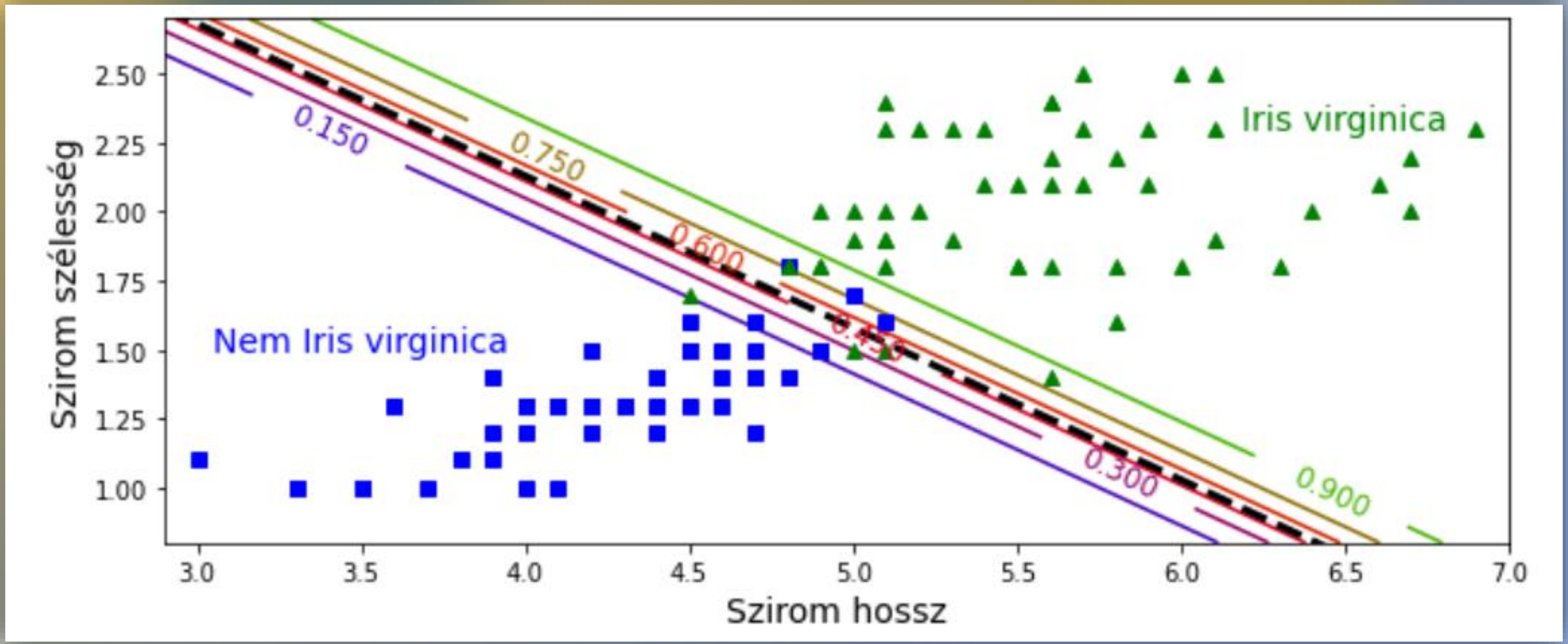
A modellünk egy 1-Dimenziós döntési határ

- Az Iris Virginica sziromszélességei 1.4-től $2.5cm$-ig terjednek, míg a többi Iris virág szirmai 0.1 és $1.8cm$ közöttiek.
- Vegyük észre, hogy van egy pár elemnyi átfedés a két csoport között.
- $2cm$ fölött a modell egészen biztos benne, hogy Virginicáról van szó, $1cm$ alatt szinte biztos benne, hogy nem tartozik az osztályba.
  A két érték között viszont bizonytalan.



2-Dimenziós döntési határok

- Egy változó (sziromszélesség) alapján 1-D döntési határt kapunk.
- Ha bevesszük a modellbe a sziromhossz változót is, a logisztikus regresszió segítségével létrehozható egy 2-D döntési határ.
- A határ felett a pozitív osztály, alatta a negatív osztály található.
- Érdemes megfigyelni, hogy még mindig vannak átfedések.



Softmax regresszió

- A logisztikus regresszió általánosítható olyan módon, hogy több osztályt is képes legyen közvetlenül megjósolni.
  Ez a **softmax**, vagy multinomiális logisztikus regresszió.
- Az algoritmus: ha van egy x egyed, a softmax modell kiszámítja $s_k(x)$-et minden k osztályra, az oztály θ paraméter vektorának segítségével.
- Egy k osztály softmax értéke: $s_k(\mathbf{x}) = \mathbf{x}^T \boldsymbol{\theta}^{(k)}$
- Majd minden osztályra kiszámítja a beletartozás valószínűségét (p̂), a softmax (normalizált exponenciális) függvény segítségével azáltal, hogy s_k-t behelyettesíti a softmax függvénybe:

$$\hat{p}_k = \sigma(\mathbf{s}(\mathbf{x}))_k = \frac{\exp\left(s_k(\mathbf{x})\right)}{\sum_{j=1}^{K} \exp\left(s_j(\mathbf{x})\right)}$$

Softmax predikciók

- Miután minden osztályra kiszámolta a softmax modell, hogy egy adott x mintaegyed mekkora valószínűséggel tartozik k osztályba, az algoritmus kiválasztja azt az osztályt, amihez a legnagyobb valószínűség tartozik:

$$\hat{y} = \underset{k}{\operatorname{argmax}}\ \sigma(\mathbf{s}(\mathbf{x}))_k = \underset{k}{\operatorname{argmax}}\ s_k(\mathbf{x}) = \underset{k}{\operatorname{argmax}}\ \left(\left(\boldsymbol{\theta}^{(k)}\right)^T \mathbf{x}\right)$$

- Az argmax operátor a változónak azt az értékét téríti vissza, amelyik maximalizál egy függvényt.
- Ebben a példában k-nak azt az értékét, amelyikhez a legnagyobb becsült valószínűség tartozik: $\sigma(s(x))_k$

A softmax modell

- A kép a létrejövő döntési határokat mutatja. Vegyük észre, hogy a három osztály között a döntési határok lineárisak.
- Az ábrán a görbe vonalakkal látható a versicolor-hoz tartozó valószínűség.
- Képes-e a modell 50% alatti valószínűséggel osztályozni?
- Mekkora valószínűség tartozik a három határ találkozásához?