Gépi tanulás a gyakorlatban

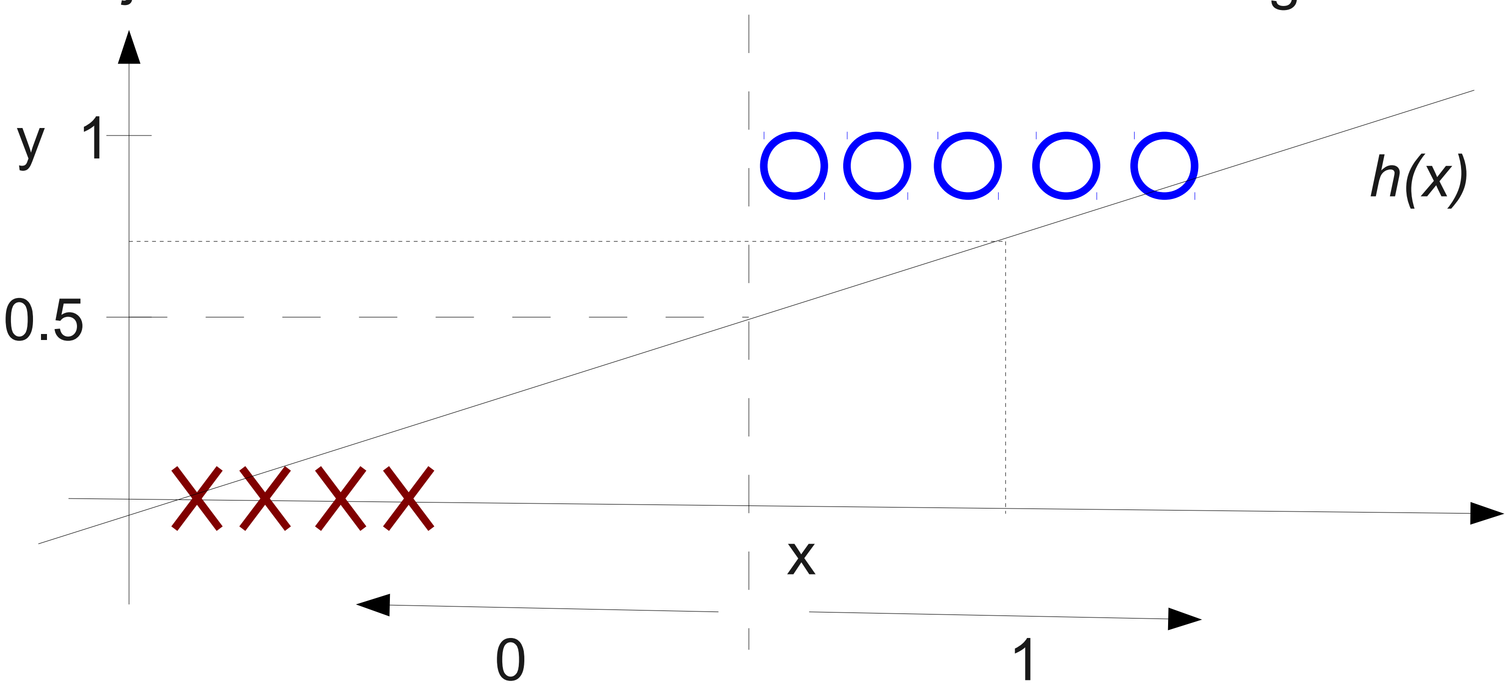
Logisztikus regresszió, túltanulás

Osztályozás – Logisztikus Regresszió

- Legyen adott egy osztályozási adatbázis:
  - x – tumor mérete,
  - y – rosszindulatú-e a daganat (Igen/Nem).

Szeretnénk *a rosszindulatúságot* előrejelezni.

- <u>Ötlet:</u>
  - Tanuljunk lineáris modell-t a szeparációra
  - Alkalmazzuk a $0.5 < h(x) \to 1$, különben 0 szabályt
- <u>Probléma:</u>
  - Ebben az esetben *h(x)* tetszőleges lehet...

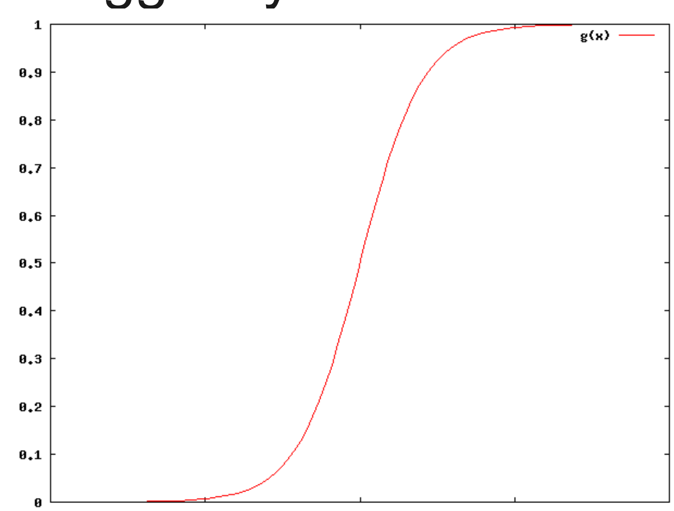


Osztályozás – Logisztikus Regresszió

- Képezzük le a *h(x)* kimenetet a [0;1] intervallumba egy *g(x)* függvény segítségével, ami:
  - 0.5-öt ad, ha az „egyenesen vagyunk” (decision boundary),
  - h(x) < 0.5-nél ha „alatta”,
  - h(x) > 0.5-nél pedig „felette”.
  - Monoton és (parciálisan) deriválható (Azért, hogy egy alkalmas költségfüggvény választása esetén tudjuk alkalmazni a gradiens módszert.)
- Jelölt *g(x)* szerepére: Sigmoid vagy Logisztikus függvény:

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

$$h_\Theta^{logreg}(x) = g(h_\Theta^{linreg}(x)) = \frac{1}{1+e^{-h_\Theta(x)}} = \frac{1}{1+e^{-\Theta^T x}}$$



- <u>Tanulás:</u> Paraméter illesztés a tanító halmazra
- <u>Interpretáció:</u> P(y=1|Theta;x) <u>Osztályozás:</u> y* = argmax_y P(y|Theta;x)

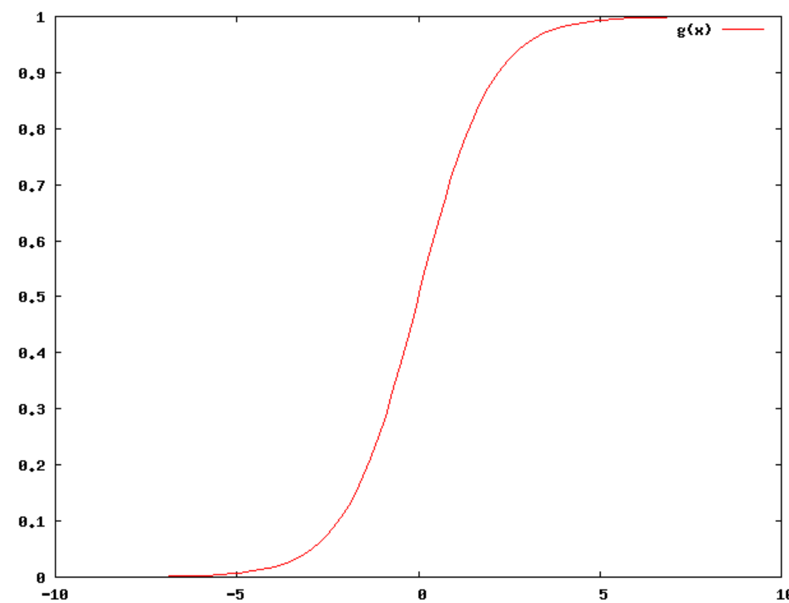
Osztályozás – Logisztikus Regresszió

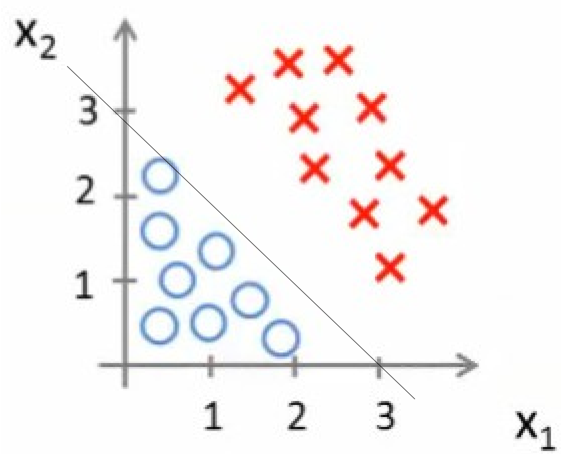
<u>1. Kvíz:</u> Tegyük fel, hogy egy *bináris osztályozási probléma* esetén betanított logreg kimenete egy x példára h(x)=P(y=1|Theta;x)=0.7. Mi a P(y=0|Theta;x) becslése:

- 1. P(y=0|Theta;x) = 0.7
- 2. P(y=0|Theta;x) = 0.7 * 0.7
- 3. P(y=0|Theta;x) = 0.3
- 4. P(y=0|Theta;x) = 0.3 * 0.7

Osztályozás – Logisztikus Regresszió

- Predikció:
  - *y=1*, ha *h(x)≥0.5* azaz *z≥0*
  - *y=0, ha h(x)<0.5 azaz z<0*
- Döntési felület: sík
  - Két térrész
  - Folytonos átmenet





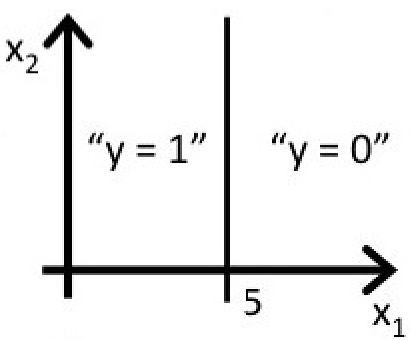
$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$$

Theta:=[-3,1,1]' ← tegyük fel, ezt tanultuk

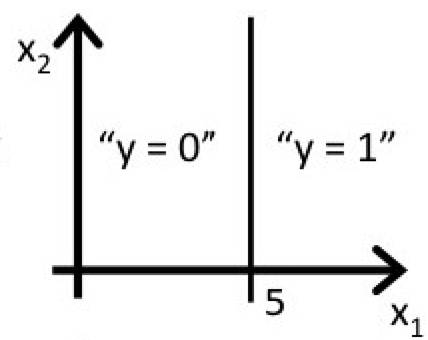
Osztályozás – Logisztikus Regresszió

<u>2. Kvíz:</u> Legyen adott egy két dimenziós (x1,x2) bementi tér, ezen egy tanuló algoritmussal a Theta0=5, Theta1=-1, Theta2=0 paramétereket tanultuk (Azaz h(x)=g(5-x1)). Melyik a helyes döntési felület?

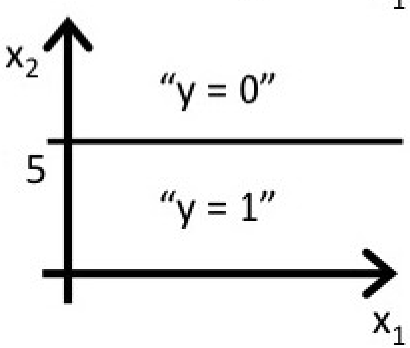
1.



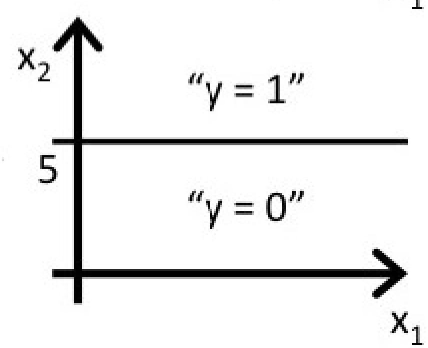
2.



3.



4.

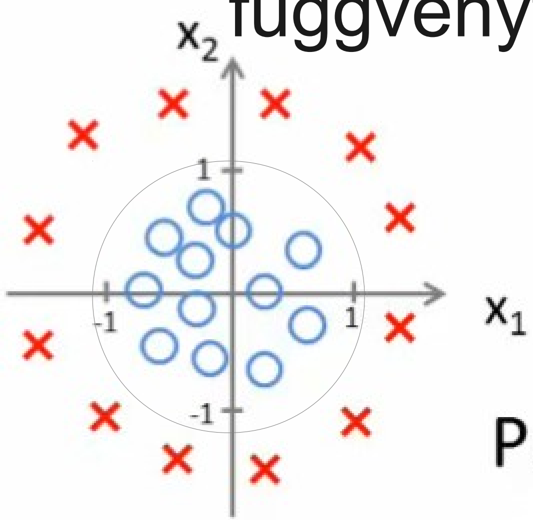


Osztályozás – Logisztikus Regresszió

- Nemlineáris döntési felület:
  - Lineáris regresszió helyett, polinomiális regresszió → nagyobb fokú tagok bevétele + gradiens módszer
    - Több paraméter (Theta nagyobb dimenziós)
    - Nagyobb reprezentációs erő (Változatosabb a felületek)
  - Logisztikus regresszió esetén: Erre polinomiális regresszió kimenetére alkalmazzuk a sigmoid függvényt → ábrán kör alakú döntési felület

Túl nagy komplexitás: túltanulás → probléma!!! (lásd később)

Általában radial basis function alapú megközelítés



$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2)$$

Theta:=[-1,0,0,1,1]' ← tegyük fel, ezt tanultuk

Predict "$y = 1$" if $-1 + x_1^2 + x_2^2 \geq 0$

Logisztikus Regresszió - Tanítás

Legyen adott egy m elemű tanulóhalmaz:

$$\mathcal{S} = \{(x^{(1)}, y^{(1)}), \ldots, (x^{(m)}, y^{(m)})\} \subseteq \mathbb{R}^{d+1} \times \{0,1\}, \forall i : x_0^{(i)} = 1$$

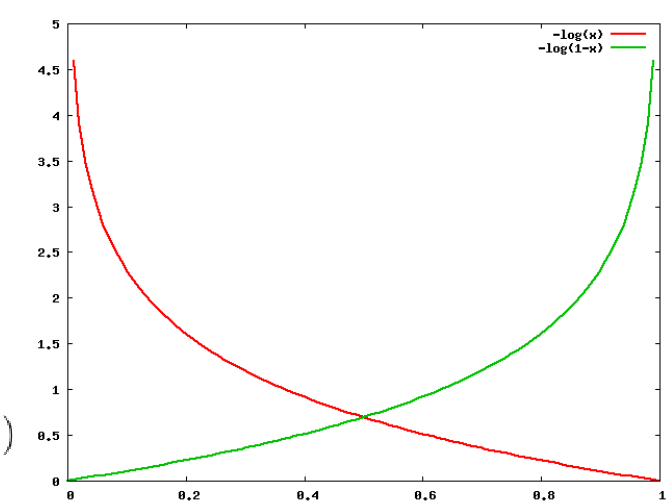
$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}}$$

Lineáris regresszió költségfüggvénye:

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \underbrace{\frac{1}{2}(\theta^T x^{(i)} - y^{(i)})^2}_{\text{Modell } h_\theta(x^{(i)}) \text{ és az elvárt érték } y^{(i)} \text{ hibája.}} = \frac{1}{m} \sum_{i=1}^{m} cost(h_\theta(x^{(i)}), y^{(i)})$$

Ahol $cost(h_\theta(x), y) = \frac{1}{2}(h_\theta(x), y)^2$. Ez a lineáris regresszió sesetén parabólikus (kvadratikus) $\to$ θ-ban konvex!
A Logisztikus regresszió esetén, ahol $h_\theta(x) = \frac{1}{1+e^{-h_\Theta(x)}} = \frac{1}{1+e^{-\Theta^T x}}$ a *cost* nem konvex és így $J(\theta)$ sem az $\to$ Gradiens módszer lokális optimumban maradhat!
Ötlet olyan *cost* megvalósítás használjuk, ami mellett $J(\theta)$ konvex!
Legyen:

$$cost(h_\theta(x), y) = \begin{cases} -log(h_\theta(x)) & \text{ha } y = 1 \\ -log(1 - h_\theta(x)) & \text{ha } y = 0 \end{cases}$$



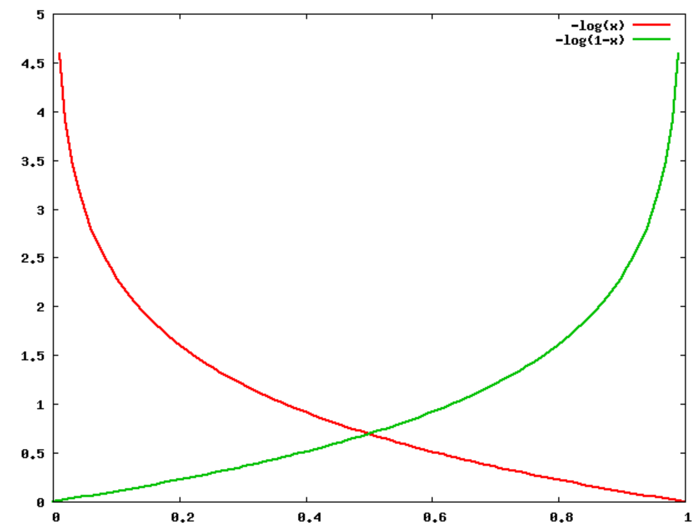
ekkor $J(\theta)$-t felírhatjuk a következő módon:

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} (y^{(i)} log(h_\theta(x)) + (1 - y^{(i)}) log(1 - h_\theta(x))$$

Osztályozás – Logisztikus Regresszió

Könnyen látható, hogy:

$$cost(h_\theta(x), y)\begin{cases} = 0 & \text{ha } y = 1 \text{ és } h_\theta(x) = 1 \\ \to \infty & \text{ha } y = 1 \text{ és } h_\theta(x) \to 0 \\ = 0 & \text{ha } y = 0 \text{ és } h_\theta(x) = 0 \\ \to \infty & \text{ha } y = 0 \text{ és } h_\theta(x) \to 1 \end{cases}$$



Ahhoz, hogy Gradiens módszer használatával tanuljuk a paramétereket, szükség lesz a $J(\theta)$ parciális deriváltjára.
Számoljuk ki:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}(y^{(i)}log(h_\theta(x)) + (1-y^{(i)})log(1-h_\theta(x))), \text{ ahol } h_\theta(x) = \frac{1}{1+e^{-\theta^T x}} \text{ és}$$

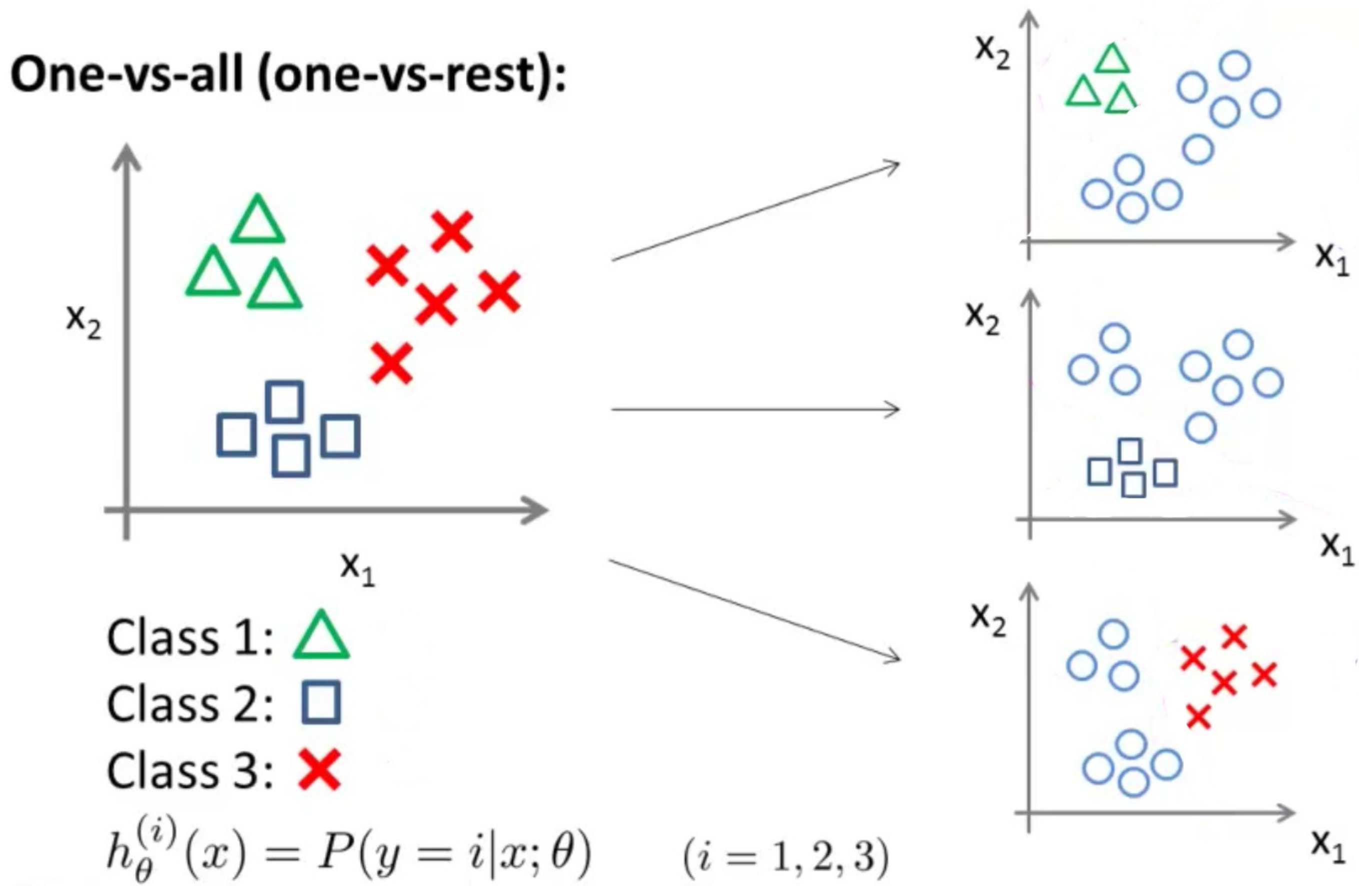
$$\frac{\partial}{\partial\theta}h_\theta(x) = \frac{e^{-\theta^T x}x}{(1+e^{-\theta^T x})^2} = \frac{1}{1+e^{-\theta^T x}} \cdot \frac{1+e^{-\theta^T x}-1}{1+e^{-\theta^T x}} \cdot x = h_\theta(x)(1-h_\theta(x))x$$

$$\frac{\partial}{\partial\theta}J(\theta^{(t)}) = \frac{\partial}{\partial\theta}J(\theta^{(t)}) = -\frac{1}{m}\sum_{i=1}^{m}(\frac{y^{(i)}}{h_{\theta^{(t)}}(x^{(i)})}\frac{\partial}{\partial\theta}h_{\theta^{(t)}}(x^{(i)}) + \frac{(1-y^{(i)})}{1-h_{\theta^{(t)}}(x^{(i)})}(-\frac{\partial}{\partial\theta}h_{\theta^{(t)}}(x^{(i)}))) =$$

$$= -\frac{1}{m}\sum_{i=1}^{m}(y^{(i)}(1-h_{\theta^{(t)}}(x^{(i)}))x^{(i)}-(1-y^{(i)})h_{\theta^{(t)}}(x^{(i)})x^{(i)}) = -\frac{1}{m}\sum_{i=1}^{m}(y^{(i)}-h_{\theta^{(t)}}(x^{(i)}))x^{(i)} = \frac{1}{m}\sum_{i=1}^{m}(h_{\theta^{(t)}}(x^{(i)})-y^{(i)})x^{(i)}$$

Több osztályos osztályozás

- Több osztályos problémák:
  - Email-ek mappákba sorolása: barátok, család, …
  - Orvosi diagnózis: megfázás, tüdőgyulladás, …
- One-vs-all

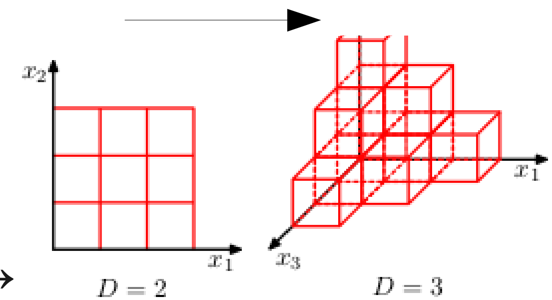


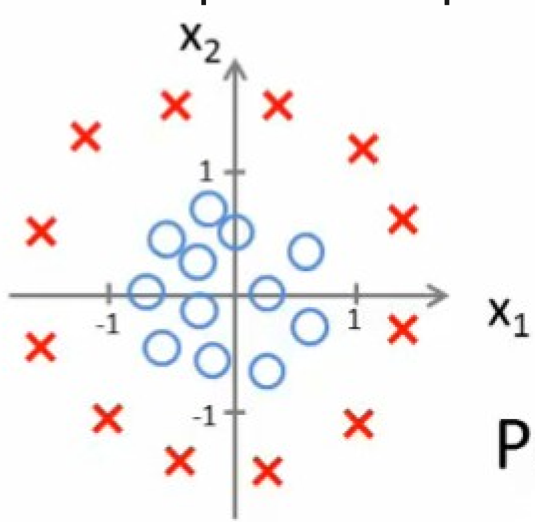
Radial basis function

Radial basis function model

Egy a tanító példák $x^{(i)}$-k helyett használjuk a tanítás és a kiértékelés során a $\phi(x^{(i)})$-ket, ahol $\phi : \mathbb{R}^d \to \mathbb{R}^n$ nemlineáris leképezés.

- Előny:
  - Nem lineáris döntési felület érhető el
- Hátrány:
  - Nagyobb dimenzióban csökken a példa sűrűség → több példa kell
  - Komplexebb hipotézis → túltanulás veszélye



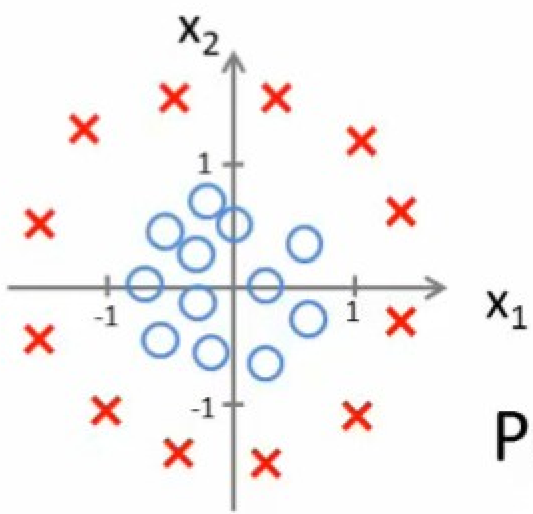


$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2)$$

Predict “$y = 1$“ if $-1 + x_1^2 + x_2^2 \geq 0$

Osztályozás – Logisztikus Regresszió

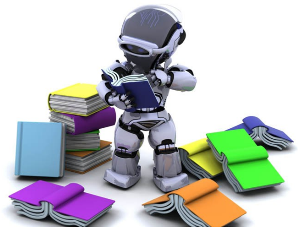
<u>3. Kvíz:</u> Milyen jellemző leképezést (radial basis function-t) használtunk az előző példában:



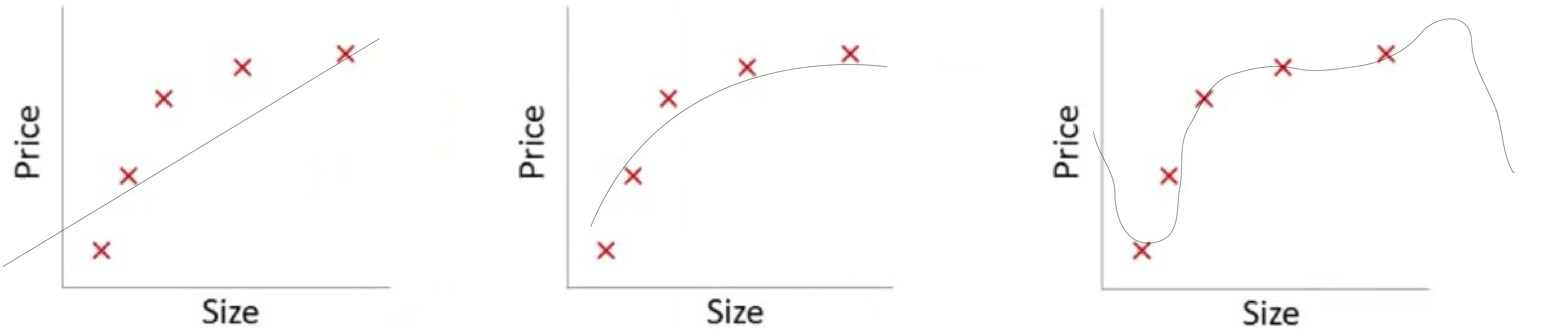
$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2)$$

Predict "$y = 1$" if $-1 + x_1^2 + x_2^2 \geq 0$

1. Phi(x)=[0, x1, x2, x1*x1, x2*x2]'
2. Phi(x)=[1, x1, x2, x1*x1, x2*x2]'
3. Phi(x)=[1, x1, x2, x1*x2, x2*x1]'
4. Nincs jellemző leképezés



Túltanulás



$\theta_0 + \theta_1 x$

$\theta_0 + \theta_1 x + \theta_2 x^2$

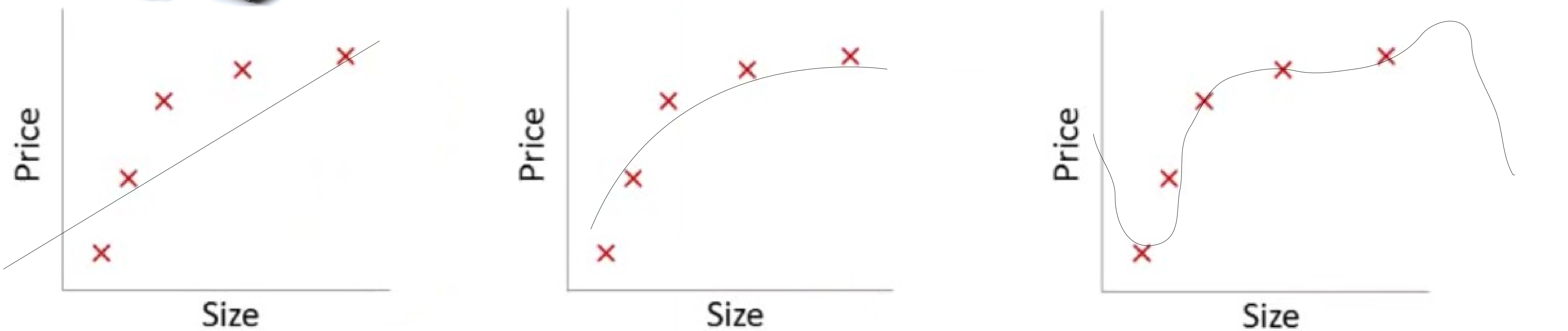
$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$

- Alultanulás (Underfit):
  - A modell reprezentációs ereje túl kicsi, nem tud jól ráilleszkedni a tanulópéldákra
- Túltanulás (Overfit):
  - A modell reprezentációs ereje túl nagy, túl jól illeszkedik a tanuló példákra → nem tud jól általánosítani, azaz új, a tanítás során nem látott példákon nem működik jól a modell.
    - Túl sok jellemző

Túltanulás elkerülése

- Plot :)   (← csak kis dimenzióban működik)
- Csökkentsük a jellemzők számát és nézzük meg, hogy tanulható marad-e az probléma.
- Modell szelekciós algoritmusok használata
- Regularizáció:
  - Tartsuk meg a jellemzőket, viszont *büntessük* a tanulás során azt, ha túl nagy értékeket vesznek fel a jellemzőkhöz rendelt paraméterek

Regularizáció



$$\theta_0 + \theta_1 x \qquad \theta_0 + \theta_1 x + \theta_2 x^2 \qquad \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x$$

- Ötlet büntessük a nagy fokú tagokat!

Egy a tanító példák $x^{(i)}$-k helyett használjuk a tanítás és a kiértékelés során a $\phi(x^{(i)})$-ket, ahol $\phi : \mathbb{R}^d \to \mathbb{R}^n$ nemlineáris leképezés.
Regularizáció alkalmazása során büntetjük a hipotézistér nagy reprezentációs erejét elősegítő tagokat.

$$min_\theta \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2 + \underbrace{100 \cdot \theta_3^2 + 100 \cdot \theta_4^2}_{\text{Regularizációs tag}}$$

Regularizáció

Működés: kis értékű paraméterek $\rightarrow$ egyszerűbb hipotézis $\rightarrow$ kisebb esély a túlillesztésre
Túl sok jellemző esetén nem tudjuk, hogy melyik nem fontos $\rightarrow$ regularizáljuk mindet (kivéve a 0-dikat (bias))
Álatalában (Lineáris regresszió)

$$J(\theta) = \frac{1}{2m}\left[\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 + \lambda\sum_{j=1}^{d}\theta_j^2\right]$$

(Logisztikus regresszió)

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}(y^{(i)}log(h_\theta(x)) + (1-y^{(i)})log(1-h_\theta(x))) + \frac{\lambda}{2m}\sum_{j=1}^{d}\theta_j^2$$

λ regularizációs paraméter, trade-off a hiba minimalizálás és a modell egyszerűn tartása között.
Túl nagy $\rightarrow$ csak a bias tag marad $\rightarrow$ csak egy horizontális egyenes marad
Túl kicsi $\rightarrow$ nincs regularizáció $\rightarrow$ túlillesztés veszélye
Gradiensek (Lineáris regresszió)

$$\frac{1}{m}\sum_{i=1}^{m}(h_{\theta^{(t)}}(x^{(i)}) - y^{(i)})x^{(i)} + \frac{\lambda}{m}\theta'^{(t)}, \text{ ahol } \theta_0'^{(t)} = 0!!!$$

Gradiens update változása (Lineáris regresszió és Logisztikus regresszió)

$$\theta_0^{(t+1)} = \theta_0^{(t)} - \alpha\frac{\partial}{\partial\theta}J(\theta^{(t)})_0 = \theta_0^{(t)} - \alpha\frac{1}{m}\sum_{i=1}^{m}(h_{\theta^{(t)}}(x^{(i)}) - y^{(i)})x_0^{(i)}$$

$$\forall 1 \leq j \leq d$$

$$\theta_j^{(t+1)} = \theta_j^{(t)} - \alpha\frac{\partial}{\partial\theta}J(\theta^{(t)})_j = (1-\alpha\frac{\lambda}{m})\theta_j^{(t)} - \alpha\frac{1}{m}\sum_{i=1}^{m}(h_{\theta^{(t)}}(x^{(i)}) - y^{(i)})x_j^{(i)}$$

Lineáris regresszió – Weka megvalósítás

- Classifier osztály kiterjesztése
- void buildClassifier(Instances) ← tanulás megvalósítása
- double classify(instance) ← predikció megvalósítása