*MODEL JEDNOSTAVNE LINEARNE REGRESIJE (JEDNA ULAZNA VARIJABLA)

m - BROJ PODATAKA

x - ULAZNA VARIJABLA

y - IZLAZNA VARIJABLA

MODEL FUNKCIJA

- CILJ: PRONAĆI FUNKCIJU $h_\theta : \mathbb{R} \to \mathbb{R}$, $h_\theta(x) = \theta_0 + \theta_1 x$ T.D. $J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(y^{(i)} - h_\theta(x^{(i)})\right)^2 \to \min_\theta$
- PSEUDOKOD GRADIJENTNE METODE:
  1. POSTAVI $\theta^{(0)}$ NA POČETNU VRIJEDNOST
  2. ODABERI PROIZVOLJNU DULJINU KORAKA $\alpha > 0$
  3. POSTAVI $k = 0$
  4. PONAVLJAJ DO KONVERGENCIJE (TJ. KADA $\|\nabla J(\theta^{(k)})\| < \epsilon$ ILI $k \geq k_{max}$)
  $$\theta^{(k+1)} \leftarrow \theta^{(k)} - \alpha \nabla J(\theta^{(k)})$$
  $$k \leftarrow k+1$$
  5. VRATI $\theta^{(k)}$
- $\nabla J(\theta^{(k)}) = \frac{1}{m} \sum_{i=1}^{m} \left(y^{(i)} - \theta^T x^{(i)}\right) x_k^{(i)}$ = U KODU = $(X\theta - y)^T X$

*MODEL VIŠESTRUKE LINEARNE REGRESIJE (VIŠE ULAZNIH VARIJABLI)

n - BROJ VARIJABLI

m - BROJ PODATAKA

$X = [x_j^{(i)}]_{m \times n}$ MATRICA DIZAJNA

$y = [y^{(i)}]_{m \times 1}$ IZLAZ

- MODEL FUNKCIJE: $h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n = \theta^T x$

$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}$ - PARAMETRI HIPERRAVNINE $\quad x = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ - ULAZ

- OPTIMIZACIJSKI PROBLEM $J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(y^{(i)} - h_\theta(x^{(i)})\right)^2 \to \min_\theta$

*POLINOMNA REGRESIJA

- MODEL FUNKCIJU IZ POLINOMNE REGRESIJE $h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \dots \theta_n x^n$

ZAPIŠEMO U KONTEKSTU LINEARNE REGRESIJE $h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \dots + \theta_n x_n$

GDJE JE $x = x_1$, $x^2 = x_2$, $x^3 = x_3$, ..., $x^n = x_n$

*FEATURE SCALING

- UGRAĐENE METODE KAO STANDARDSCALER ILI MINMAXSCALER

*ODREĐIVANJE PARAMETARA θ POMOĆU SUSTAVA NORMALNIH JEDNADŽBI

$$\theta = (X^T X)^{-1} X^T y$$

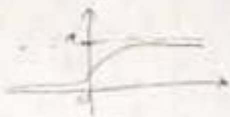
- AKO $X^T X$ NIJE REGULARNA TJ. NEMA INVERZ KORISTIMO PSEUDOINVERZ

* METODA POTPORNIH VEKTORA (SVM)

- CILJ JE SEPARIRATI PODATKE HIPERRAVNINOM T.D. GEOMETRIJSKA MARGINA BUDE MAKSIMIZIRANA (OVDJE JE GEOMETRIJSKA MARGINA $= \frac{1}{\|\theta\|} = \frac{1}{\text{NP. LINALG NORM(MAGNITUDE)}}$
- PODACI NE MORAJU BITI LINEARNO SEPARABILNI (UVODIMO SLACK VARIJABLE)
- OPTIMIZACIJSKI PROBLEM: $\underset{\theta, \xi_i}{\operatorname{argmin}} \frac{1}{2}\|\theta\|^2 + C\sum_{i=1}^{m} \xi_i$ UZ UVJETE $y^{(i)}\theta^T x^{(i)} \geq 1-\xi_i,\ \xi_i \geq 0,\ i=1..m$
- UGRAĐENE FUNKCIJE: - LINEARSVC ILI SVC
  - LINEARSVC JE BRŽI, ALI NE RAČUNA POTPORNE VEKTORE

* LOGISTIČKA REGRESIJA

- KLASIFIKACIJA PODATAKA MODEL FUNKCIJOM $h_\theta : \mathbb{R}^n \to \langle 0,1 \rangle$, $h_\theta(x) = \sigma(\theta^T x)$ GDJE JE $\sigma(\theta^T x) = \frac{1}{1+e^{-\theta^T x}}$ S GRAFOM



- ZA PODATAK $x^{(i)}$ MODEL FUNKCIJE VRAĆA VJEROJATNOST DA JE $x^{(i)}$ KLASE 1
- OPTIMIZACIJSKI PROBLEM: $J(\theta) = \sum_{i=1}^{m} \log(1 + e^{-y^{(i)}\theta^T x^{(i)}}) \to \min_\theta$ (ZA $y^{(i)} \in \{-1,1\}$)
- POŠTO JE $J(\theta)$ KONVEKSNA, MOŽE SE RJEŠAVATI GRADIJENTNOM METODOM
- LIKELIHOOD FUNKCIJA: $J'(\theta) = \prod_{i=1}^{m} P(y^{(i)} | \theta; x^{(i)}) \to \max_\theta$ OVOM FUNKCIJOM MAKSIMIZIRAMO VJEROJATNOST DA JE SVAKI PODATAK ISPRAVNO KLASIFICIRAN
- OPTIMIZACIJSKI PROBLEM: $J(\theta) = \sum_{i=1}^{m}\left[-y^{(i)}\log(h_\theta(x^{(i)})) - (1-y^{(i)})\log(1-h_\theta(x^{(i)}))\right]$ (ZA $y^{(i)} \in \{0,1\}$)

* STOHASTIČKA GRADIJENTNA METODA

- ODABIROM VIŠE SLUČAJNO ODABRANIH i IZ $\{1,...,m\}$ POMAKNUT ĆEMO PARAMETRE θ U SMJERU TOČNOG GRADIJENTA
- PSEUDOKOD STOHASTIČKE GRADIJENTNE METODE
  1. POSTAVI θ NA POČETNU VRIJEDNOST
  2. PONAVLJAJ DO KONVERGENCIJE
     ODABERI $i \in \{1,...,m\}$ SLUČAJNIM ODABIROM
     $\theta \leftarrow \theta - \alpha \nabla_\theta \text{loss}(h_\theta(x^{(i)}), y^{(i)})$

* MATRICA ZABUNE

	0	1
0	TN	FP
1	FN	TP

$\text{ACCURACY} = \frac{TP+TN}{SVE}$

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

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

* SOFT-MAX REGRESIJA

- GENERALIZIRATI KONCEPT LOGISTIČKE REGRESIJE ZA VIŠEKLASNU KLASIFIKACIJU T.D. JE MODEL FUNKCIJA $h_\theta : \mathbb{R}^n \to \mathbb{R}^k$, $h_\theta(x) = \begin{bmatrix} P(y=1 | x;\theta) \\ P(y=2 | x;\theta) \\ \vdots \\ P(y=k | x;\theta) \end{bmatrix} = \sigma\left(\begin{bmatrix} \theta^{(1)T}x \\ \vdots \\ \theta^{(k)T}x \end{bmatrix}\right) = \frac{1}{\sum_{j=1}^{k} e^{\theta^{(j)T}x}} \begin{bmatrix} e^{\theta^{(1)T}x} \\ \vdots \\ e^{\theta^{(k)T}x} \end{bmatrix}$
- SVOJSTVO: $\sum_{l=1}^{k} P(y=l | \theta; x) = 1$
- OPTIMIZACIJSKI PROBLEM: $J(\theta) = -\sum_{i=1}^{m}\sum_{l=1}^{k} I^{(l)}(y^{(i)}) \log\left(\frac{e^{\theta^{(l)T}x^{(i)}}}{\sum_{j=1}^{k} e^{\theta^{(j)T}x^{(i)}}}\right) \to \min_\theta$

$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}}$

→ PROBLEMI KLASIFIKACIJE

- MODEL FUNKCIJA OBLIKA $h_\theta : \mathbb{R}^n \to \{0, 1, \dots, k-1\}$
- ZA DVIJE KLASE VRIJEDI: $h_\theta(x) = \mathrm{sign}(\theta_0 + \theta_1 x_1 + \dots + \theta_n x_n) = \mathrm{sign}(\theta^T x)$
  - PODACI $\{x^{(1)}, x^{(2)}, \dots, x^{(m)}\} \subset \mathbb{R}^n$ S PRIPADNIM KLASAMA $y^{(i)} \in \{-1, 1\}$
  - PODACI $\{(x^{(i)}, y^{(i)}), i = 1, \dots, m\}$ SU LINEARNO SEPARABILNI KROZ ISHODIŠTE AKO VRIJEDI $y^{(i)} \theta^T x^{(i)} > 0 \quad \forall i = 1, \dots, m$
  - LOSS FUNKCIJA / OPTIMIZACIJSKI PROBLEM: $J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \left(1 - \delta(h_\theta(x^{(i)}), y^{(i)})\right) \to \min_\theta$

$J(\theta) = 0$ AKO SU SVI SEPARIRANI

$\delta(a, b) = \begin{cases} 1, & a = b \\ 0, & a \neq b \end{cases}$

→ PERCEPTRON ALGORITAM

- CILJ: PRONAĆI HIPERRAVNINU $\theta_0 + \theta_1 x_1 + \dots + \theta_n x_n = 0$ KOJA SEPARIRA ULAZNE PODATKE NA ONE KLASE $\underline{1}$ I KLASE $\underline{-1}$
- ZA PODATAK x KLASE 1 VRIJEDIT ĆE $\theta^T x > 0$, A KLASE -1 $\theta^T x < 0$
- VEKTOR $\vec{\theta}$ JE NORMALA HIPERRAVNINE
- AKO JE $\underline{n = 2}$, HIPERRAVNINA JE PRAVAC OPISAN JEDNADŽBOM $y = -\frac{\theta_1}{\theta_2} x - \frac{\theta_0}{\theta_2}$

IZ

$\theta_0 + \theta_1 x + \theta_2 y = 0$

- PSEUDOKOD PERCEPTRON ALGORITMA

```
POSTAVI θ^(0) = (0, ..., 0)
POSTAVI k = 0
REPEAT
        FOR i = 1 TO m
              IF y^(i) ≠ h_θ(x^(i))          θ^T x
                   θ^(k+1) ← θ^(k) + y^(i) x^(i)   PUR
                   k ← k + 1
UNTIL y^(i) θ^(k)T x^(i) > 0, ∀ i = 1, ..., m
```

- $\underline{\text{GEOMETRIJSKA MARGINA}}$ - NAJKRAĆA UDALJENOST IZMEĐU PODATKA I HIPERRAVNINE $\vec{\theta}^T x = 0$
  - OZNAKA $\gamma = \frac{\theta_0 + x\vec{\theta}}{\|\vec{\theta}\|}$
- $\underline{\text{FUNKCIJSKA MARGINA}}$ - NAJVEĆI BROJ γ ZA KOJEG VRIJEDI $y^{(i)} \theta^T x^{(i)} \geq \gamma$, $\forall i = 1, \dots, m$

* PRENAUČENOST / OVERFITTING - MODEL FUNKCIJA IMA PREVELIK BROJ PARAMETARA
  - DOBRO FITTA TRAIN, ALI LOŠE TEST PODATKE
* PODNAUČENOST / UNDERFITTING - MODEL FUNKCIJA IMA PREMAL BROJ PARAMETARA
  - LOŠE FITTA I TRAIN I TEST PODATKE

* REGULARIZACIJA
- DODAVANJE REGULARIZACIJSKE FUNKCIJE $R: \mathbb{R}^n \to \mathbb{R}$ U LOSS FUNKCIJU KAKO BI SE KONTROLIRAO UTJECAJ PARAMETARA θ U MODEL FUNKCIJI h_θ
- UVODI SE U SVRHU SPRJEČAVANJA OVERFITTINGA
- NOVA LOSS FUNKCIJA: $J(\theta) + R(\theta)$
- ZA R NPR UZMEMO L_2 NORMU I HIPERPARAMETAR λ
- ZA LINEARNU REGRESIJU DOBIJEMO LOSS FUNKCIJU: $J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (y^{(i)} - h_\theta(x^{(i)}))^2 + \lambda \sum_{j=1}^{n} \theta_j^2$
- ZA SOFT-MAX REGRESIJU -II-: $J(\theta) = -\sum_{i=1}^{m} \sum_{j=1}^{k} I\{y^{(i)}\} \log \left(\frac{e^{\theta^{(j)T} x^{(i)}}}{\sum_{l=1}^{k} e^{\theta^{(l)T} x^{(i)}}} \right) + \lambda \sum_{i=1}^{m} \sum_{j=1}^{k} \theta_{ij}^2$
- ZA REGRESIJU SA SUSTAVOM NORMALNIH JEDNADŽBI: $\theta = \left(X^T X + \lambda \begin{bmatrix} 0 & & \\ & 1 & \\ & & 1 \end{bmatrix} \right)^{-1} X^T y$

=> λ PREMAL - OVERFITTING OSTAJE OVERFITTING
=> λ PREVELIK - OVERFITTING POSTAJE UNDERFITTING

* SUBGRADIJENT
- SVAKI KOEFICIJENT SMJERA PRAVCA KOJI ZADOVOLJAVA DA PODRŽAVA HIPERRAVNINU KOJA SE U POTPUNOSTI NALAZI ISPOD KONVEKSNE FUNKCIJE
- KORISTIMO ZA KONVEKSNE FUNKCIJE KOJE IMAJU 1 ŠPIC