Linearne funkcije odlučivanja

Linearne funkcije odlučivanja su oblika:

$$d(\vec{x}) = \vec{w}_0^T \vec{x}_0 + w_{n+1} = \vec{w}^T \vec{x}$$

$$\vec{x}_0 = [\, x_1 \quad x_2 \quad \dots \quad x_n \,]^T$$

$$\vec{x} = [\, x_1 \quad x_2 \quad \dots \quad x_n \quad 1 \,]^T$$

$$\vec{w}_0 = [\, w_1 \quad w_2 \quad \dots \quad w_n \,]^T$$

$$\vec{w} = [\, w_1 \quad w_2 \quad \dots \quad w_n \quad w_{n+1} \,]^T$$

Slučaj 1. Svaki razred je od ostalih moguće odvojiti jednom hiperravninom.

$$d_i(\vec{x}) = \vec{w}_i^T \vec{x} \begin{cases} > 0 \Rightarrow \vec{x} \in \omega_i \\ \leq 0 \Rightarrow \vec{x} \notin \omega_i \end{cases}$$

Slučaj 2. Razredi su linearno razdvojivi par po par. Za M razreda je tada potrebno $M(M-1)/2$ funkcija odlučivanja.

$$d_{ij}(\vec{x}) > 0,\ \forall j \neq i,\ j = 1, \dots, M \Rightarrow \vec{x} \in \omega_i$$

Slučaj 3. Postoji M funkcija odlučivanja $d_i(\vec{x}) = \vec{w}_i^T \vec{x}$, $k = 1,\ 2, \dots,\ M$.

$$d_i(\vec{x}) > d_j(\vec{x}),\ \forall j \neq i,\ j = 1, \dots, M \Rightarrow \vec{x} \in \omega_i$$

Granice razreda su određene kao $d_i(\vec{x}) = d_j(\vec{j})$.

Gradijentni postupci

Promatramo funkciju oblika $J(\vec{w}, \vec{x})$ koja postiže ekstrem (minimum) pri ispunjavanju uvjeta:

$$\vec{w}^T \vec{x}_i > 0, \quad i = 1, 2, \dots, N$$

Pri ispitivanju gornjeg uvjeta se vektor uzorka množi s -1 ako ne pripada razredu ω karakteriziranom s vektorom $\vec{w}$.

Tada iterativno mijenjamo vrijednost vektora $\vec{w}$ prema:

$$\vec{w}_{(k+1)} = \vec{w}_{(k)} - c \left. \frac{\partial J(\vec{w}, \vec{x})}{\partial \vec{w}} \right|_{\vec{w}_{(k)}}$$

Iteracije prekidamo kada su uz dobiveni $\vec{w}$ svi uzorci pravilno razvrstani.

Algoritam perceptrona sa stalnim prirastom

$$J(\vec{w}, \vec{x}) = \frac{1}{2}\left(|\vec{w}^T \vec{x}| - \vec{w}^T \vec{x}\right)$$

$$\frac{\partial J(\vec{w}, \vec{x})}{\partial \vec{w}} = \frac{1}{2} \operatorname{sgn}(\vec{w}^T \vec{x})\vec{x} - \frac{1}{2}\vec{x}$$

$$\vec{w}_{(k+1)} = \vec{w}_{(k)} + \begin{cases} 0,\ \vec{w}_{(k)}^T \vec{x}_{(k)} > 0 \\ c\vec{x}_{(k)},\ \vec{w}_{(k)}^T \vec{x}_{(k)} \leq 0 \end{cases}$$

Algoritam perceptrona sa stalnim prirastom konvergira za linearno razdvojive razrede.

Postupak perceptrona s djelomičnom korekcijom

$$J(\vec{w}, \vec{x}) = \frac{1}{4\vec{x}^T \vec{x}}\left(|\vec{w}^T \vec{x}|^2 - |\vec{w}^T \vec{x}|\vec{w}^T \vec{x}\right)$$

$$\frac{\partial J(\vec{w}, \vec{x})}{\partial \vec{w}} = \frac{1}{2\vec{x}^T \vec{x}}\left(|\vec{w}^T \vec{x}|\vec{x} \operatorname{sgn}(\vec{w}^T \vec{x}) - |\vec{w}^T \vec{x}|\vec{x}\right)$$

$$\vec{w}_{(k+1)} = \vec{w}_{(k)} + \begin{cases} 0,\ \vec{w}_{(k)}^T \vec{x}_{(k)} > 0 \\ \lambda \dfrac{|\vec{w}_{(k)}^T \vec{x}_{(k)}|}{\vec{x}_{(k)}^T \vec{x}_{(k)}} \vec{x}_{(k)},\ \vec{w}_{(k)}^T \vec{x}_{(k)} \leq 0 \end{cases}$$

Postupak konvergira za $0 < \lambda < 2$ i $\vec{w}_{(1)} \neq \vec{0}$.

Postupak Hoa i Kashyapa (LMSE)

LMSE – *Least Mean Square Error*

$$J(\vec{w}, \vec{x}, \vec{b}) = \frac{1}{2} \sum_{j=1}^{N} (\vec{w}^T \vec{x}_j - \vec{b}_j)^2 = \frac{1}{2} \left|\left| \mathbf{X}\vec{w} - \vec{b} \right|\right|^2$$

$$\mathbf{X} = [\, \vec{x}_1^T \quad \vec{x}_2^T \quad \dots \quad \vec{x}_N^T \,]^T$$

Generalizirani inverz matrice $\mathbf{X}$

$$\mathbf{X}^{\#} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$$

$$\frac{\partial J(\vec{w}, \vec{x})}{\partial \vec{w}} = -\mathbf{X}\vec{w} + \vec{b}$$

Vektor pogreške $\vec{e}$

$$\vec{e}_{(k)} = \mathbf{X}\vec{w}_{(k)} - \vec{b}_{(k)}$$

Vektor $\vec{b}$

$$\vec{b}_{(k+1)} = \vec{b}_{(k)} + c\left(\vec{e}_{(k)} + |\vec{e}_{(k)}|\right)$$

Vektor $\vec{w}$

$$\vec{w}_{(k+1)} = \mathbf{X}^{\#} \vec{b}_{(k+1)}$$

Vektor $b_{(1)}$ mora imati sve elemente pozitivne. Postupak završava kada je $\vec{e} = 0$. Ako su sve komponente vektora pogreške $\vec{e}$ nepozitivne, ali nisu sve jednake nuli, tada razredi nisu linearno razdvojivi te možemo prekinuti postupak.

Poopćeni postupak perceptrona

Neka postoji ukupno M razreda i neka je

$$d_j(\vec{x}_{(k)}) = \vec{w}_{j\,(k)}^T \vec{x}_{(k)}, \quad j = 1, 2, \dots, M$$

U k-tom koraku promatramo uzorak $\vec{x}_{(k)} \in \omega_i$. Ako vrijedi

$$d_i(\vec{x}_{(k)}) > d_j(\vec{x}_{(k)}),\ \forall j \neq i,\ j = 1, \dots, M$$

ne mijenjamo vektor $\vec{w}_j$, tj.

$$\vec{w}_{j\,(k+1)} = \vec{w}_{j\,(k)}, \quad j = 1, \ldots, M$$

Ako vrijedi

$$\exists l,\, d_i(\vec{x}_{(k)}) \leq d_l(\vec{x}_{(k)})$$

moramo ispraviti vektore $\vec{w}_i$ i $\vec{w}_l$.

$$\vec{w}_{i\,(k+1)} = \vec{w}_{i\,(k)} + c\vec{x}_{(k)}$$

$$\vec{w}_{l\,(k+1)} = \vec{w}_{l\,(k)} - c\vec{x}_{(k)}$$

$$\vec{w}_{j\,(k+1)} = \vec{w}_{j\,(k)}, \quad j = 1, \ldots, M \wedge j \neq i, l$$

U pojedinom koraku postupka mijenjamo funkcije odlučivanja samo za jedan l, tj. ako postoji više l takvih da vrijedi

$$d_i(\vec{x}_{(k)}) \leq d_l(\vec{x}_{(k)})$$

ispravku vršimo samo za najmanji l. Postupak završavamo kada su svi vektori uzoraka ispravno razvrstani.

Poopćene funkcije odlučivanja

Umjesto vektora uzoraka koristimo uzorke s transformiranim koordinatama preko funkcija $f_i(\vec{x})$. Vektor uzoraka

$$\vec{x} = [\,x_1 \quad x_2 \quad \ldots \quad x_n \quad 1\,]^T$$

tada postaje

$$\vec{x}' = [\,f_1(\vec{x}) \quad f_2(\vec{x}) \quad \ldots \quad f_k(\vec{x}) \quad 1\,]^T.$$

Odabiremo takvu transformaciju koja za novi vektor uzoraka $\vec{x}'$ omogućuje korištenje linearnih funkcija odlučivanja.

Kvadratna funkcija. Funkcija odlučivanja postaje

$$d(\vec{x}) = \sum_{j=1}^{n} w_{jj} x_j^2 + \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} w_{jk} x_j x_k + \sum_{j=1}^{n} w_j x_j + w_{n+1}$$

Postupak učenja s potencijalnim funkcijama

Potencijalna funkcija mora zadovoljavati sljedeća svojstva:

1. Funkcija je simetrična, $f(\vec{x}, \vec{x}_{(k)}) = f(\vec{x}_{(k)}, \vec{x})$.
2. Funkcija trne kada $\vec{x}$ teži prema beskonačnosti, $||\vec{x} - \vec{x}_{(k)}|| \to \infty \Rightarrow f(\vec{x}, \vec{x}_{(k)}) \to 0$.
3. Funkcija ima ekstrem za $\vec{x} = \vec{x}_{(k)}$.

Ako je u k-tom koraku uzorak pogrešno razvrstan mijenjamo funkciju odlučivanja:

$$d_{(k+1)}(\vec{x}) = d_{(k)}(\vec{x}) + r_{(k)} f(\vec{x}, \vec{x}_{(k)})$$

$$r_{(k)} = \begin{cases} 0, \, \vec{x}_{(k)} \in \omega \wedge d_{(k)}(\vec{x}_{(k)}) > 0 \\ 0, \, \vec{x}_{(k)} \notin \omega \wedge d_{(k)}(\vec{x}_{(k)}) < 0 \\ 1, \, \vec{x}_{(k)} \in \omega \wedge d_{(k)}(\vec{x}_{(k)}) \leq 0 \\ -1, \, \vec{x}_{(k)} \notin \omega \wedge d_{(k)}(\vec{x}_{(k)}) \geq 0 \end{cases}$$

Postupak dopušta učenje bez pretpostavki o konačnom obliku funkcije odlučivanja.

Klasifikacija pomoću udaljenosti

Za klasifikaciju pomoću udaljenosti moramo definirati razdaljinsku funkciju $d : X^n \to \mathbb{R}_+$ koja zadovoljava sljedeća svojstva:

1. $\forall \vec{x}, \vec{y} \in X : \; d(\vec{x}, \vec{y}) \geq 0$
2. $d(\vec{x}, \vec{y}) = 0 \Leftrightarrow \vec{x} = \vec{y}$
3. $\forall \vec{x}, \vec{y} \in X : \; d(\vec{x}, \vec{y}) = d(\vec{y}, \vec{x})$
4. $\forall \vec{x}, \vec{y}, \vec{z} : \; d(\vec{x}, \vec{y}) \leq d(\vec{x}, \vec{z}) + d(\vec{z}, \vec{y})$

Euklidova udaljenost

$$d(\vec{x}, \vec{y}) = ||\vec{x} - \vec{y}|| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

Udaljenost Minkowskog

$$d(\vec{x}, \vec{y}) = \left(\sum_{i=1}^{n} |x_i - y_i|^s \right)^{1/s}$$

Za $s = 2$ udaljenost postaje Euklidska, za $s = 1$ udaljenost postaje "city block" udaljenost (Manhattan), a za $s \to \infty$ udaljenost postaje Čebiševljeva udaljenost.

Klasifikacija na temelju jednog prototipa. Svaki razred ω_i je predstavljen tipičnim predstavnikom $\vec{z}_i$.

$$\forall j \neq i, j = 1, \ldots, M : \; d(\vec{x}, \vec{z}_i) < d(\vec{x}, \vec{z}_j) \Rightarrow \vec{x} \in \omega_i$$

Klasifikacija na temelju više prototipova. Svaki razred ω_i je predstavljen s više predstavnika (prototipova) $\vec{z}_i^{\,1}$, $\vec{z}_i^{\,2}$,..., $\vec{z}_i^{\,N_i}$.

$$D(\vec{x}, \omega_i) = \min_k d(\vec{x}, \vec{z}_i^{\,k})$$

$$\forall j \neq i, j = 1, \ldots, M : \; D(\vec{x}, \omega_i) < D(\vec{x}, \omega_j) \Rightarrow \vec{x} \in \omega_i$$

$$d_i(\vec{x}) = \max_k \left(\vec{x}^T \vec{z}_i^{\,k} - \frac{1}{2} \vec{z}_i^{\,k\,T} \vec{z}_i^{\,k} \right)$$

Klasifikacijsko pravilo najbližeg susjeda (NN)

NN – *Nearest Neighbour*

Uzorak $\vec{x}$ razvrstavamo u razred ω_i ako je njemu najbliži susjed iz razreda ω_i.

1-NN algoritam

Vjerojatnost pogreške

$$p_e = \frac{1}{2^N}$$

q-NN algoritam

Vjerojatnost pogreške

$$p_e = \frac{1}{2^N} \sum_{\alpha=0}^{(q-1)/2} C_\alpha^N$$

$$\alpha \leq (q-1)/2$$

Za q-NN algoritam pogreška se događa ako je broj uzoraka koji pripada ispravnom razredu manji od polovine broja susjeda koje promatramo.

Grupiranje

Razlikujemo heurističke metode i metode koje se temelje na minimizaciji (ili maksimizaciji) neke kriterijske funkcije.

Jednostavan heuristički postupak

Promatramo skup od N uzoraka.

1. Nasumce odaberemo središte prve grupe $\vec{z}_1$ i nenegativni prag T.
2. Računamo udaljenost sljedećeg uzorka od središta grupe i ako je veća od praga tvorimo novo središte $\vec{z}_2$. Inače dodjeljujemo uzorak postojećoj grupi.
3. Za svaki od sljedeći uzoraka računamo sve udaljenosti od postojećih središta grupa, i ako su sve veće od praga stvaramo novo središte, a inače dodjeljujemo uzorak grupi čijem središtu je najbliži.

Algoritam min-max udaljenosti

Promatramo skup od N uzoraka.

1. Odaberemo središte prve grupe $\vec{z}_1$.
2. Pronađemo najudaljeniji uzorak od prvog središta $\vec{z}_1$ i stvorimo novu grupu s njim kao središtem (označimo ga s $\vec{z}_2$).
3. Računamo udaljenosti svih preostalih uzoraka od oba središta. Za svaki uzorak od te dvije udaljenosti odaberemo manju, a iz tako dobivenog skupa odaberemo maksimalnu udaljenost.
4. Ako je udaljenost dobivena u prethodnom koraku usporediva s udaljenosti između $\vec{z}_1$ i $\vec{z}_2$ stvaramo novu grupu s uzorkom koji odgovara toj udaljenosti kao središtem, inače algoritam završava.
5. Ponavljamo određivanje maksimuma skupa minimalnih udaljenosti i stvaranje novih grupa sve dok udaljenost prestane biti usporediva s udaljenošću $d(\vec{z}_1, \vec{z}_2)$. Sve preostale uzorke rasporedimo prema najmanjoj udaljenosti.

Algoritam K srednjih vrijednosti

Promatramo skup od N uzoraka.

1. Odaberemo $K < N$ središta grupa. Neka su to $\vec{z}_{1\,(1)}, \vec{z}_{2\,(1)},\ldots,\vec{z}_{K\,(1)}$.
2. U svakom koraku razdijelimo uzorke u grupe, i to $\vec{x}$ u grupu $S_{i\,(k)}$ ako je

   $$||\vec{x} - \vec{z}_{i\,(k)}|| < ||\vec{x} - \vec{z}_{j\,(k)}||,\ j = 1, \ldots, K \wedge i \neq j.$$

   $S_{i\,(k)}$ je grupa čije je središte $\vec{z}_{i\,(k)}$
3. Izračunamo nova središta grupa prema

   $$\vec{z}_{i\,(k+1)} = \frac{1}{\operatorname{card} S_{i\,(k)}} \sum\nolimits_{\vec{x} \in S_{i\,(k)}} \vec{x}.$$

   Ovako određena središta grupa minimiziraju kriterijsku funkciju

   $$J = \sum\nolimits_{i=1}^{K} \sum\nolimits_{\vec{x} \in S_{i\,(k)}} ||\vec{x} - \vec{z}_{i\,(k)}||^2.$$

4. Postupak završava kada je

   $$\vec{z}_{i\,(k+1)} = \vec{z}_{i\,(k)},\ i = 1, \ldots, K$$

Grupiranje na temelju teorije grafova

Matrica sličnosti. Potrebno je odrediti matricu sličnosti. Elementi matrice sličnosti su

$$s_{kl} = \begin{cases} 1, \, D_{kl} \leq \theta \\ 0, \, D_{kl} > \theta \end{cases},$$

gdje je

$$D_{kl} = ||\vec{x}_k - \vec{x}_l||, \quad k, l = 1, 2, \ldots, N,$$

a θ je vrijednost praga. Postupak:

1. Odabiremo redak s najviše “1”, npr. neka je to i-ti redak (odgovara i-tom elementu).
2. Promatramo stupce u i-tom retku koji također sadrže “1”, npr. j, k i l. Promatramo retke od j, k i l koji također sadrže “1” (povezani su s i). Ponavljamo postupak dok ne pokupimo sve elemente koji su povezani s i-tim elementom.
3. Brišemo elemente koji su povezani s i-tim elementom (retke i stupce i, j, k,...), te prelazimo na sljedeći element.
4. Postupak ponavljamo dok ne uklonimo sve retke.

Stablo minimalne duljine. Rezanjem grana u stablu minimalne duljine dobivamo grupe. Grane koje treba prerezati možemo odrediti prema histogramu udaljenosti.

Statistički pristup raspoznavanju uzoraka

Vjerojatnost unije

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Vjerojatnost presjeka

$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$$

Formula potpune vjerojatnosti

$$P(B) = \sum\nolimits_{i=1}^{n} P(A_i)P(B|A_i)$$

Bayesova formula

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{n} P(A_j)P(B|A_j)}$$

Bayesov klasifikator

$d_i(\vec{x}) = p(\omega_i|\vec{x})p(\vec{x})$

Bayesov klasifikator za uzorke s normalnom razdiobom

Jednodimenzionalni slučaj

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right)$$

$$m = \mathrm{E}[x] = \int_{-\infty}^{\infty} x\, p(x)dx$$

$$\sigma^2 = \mathrm{E}\big[(x-m)^2\big] = \int_{-\infty}^{\infty} (x-m)^2\, p(x)dx$$

Višedimenzionalni slučaj

$$p(\vec{x}) = \frac{1}{\sqrt{(2\pi)^n|\mathbf{C}|}}\exp\Big(-\frac{1}{2}(\vec{x}-\vec{m})^T\mathbf{C}^{-1}(\vec{x}-\vec{m})\Big)$$

$$\vec{m} = \mathrm{E}[\vec{x}]$$

$$\mathbf{C} = \mathrm{E}\big[(\vec{x}-\vec{m})(\vec{x}-\vec{m})^T\big]$$

$$\mathbf{C}_{ij} = \mathrm{E}\big[(x_i-m_i)(x_j-m_j)\big]$$

Procjena parametara razdiobe

$$\vec{m} = \mathrm{E}[\vec{x}] \approx \frac{1}{N}\sum_{i=1}^{N}\vec{x}_i$$

$$\mathbf{C} = \mathrm{E}\big[(\vec{x}-\vec{m})(\vec{x}-\vec{m})^T\big] \approx \frac{1}{n}\sum_{i=1}^{N}\vec{x}_i\vec{x}_i^T - \vec{m}\vec{m}^T$$

Umjesto funkcije $d_i(\vec{x})$ promatramo njen logaritam.

$$\ln d_i(\vec{x}) = \ln p(\omega_i) - \frac{1}{2}\ln|\mathbf{C}_i| - \frac{1}{2}(\vec{x}-\vec{m}_i)^T\mathbf{C}_i^{-1}(\vec{x}-\vec{m}_i)$$

Posebni slučajevi:

a) Ako je $\mathbf{C}_i = \mathbf{C}$ (disperzijska matrica je jednaka za sve razrede) funkcija odlučivanja postaje:

$$d_i(\vec{x}) = \ln p(\omega_i) + \vec{x}^T\mathbf{C}^{-1}\vec{m}_i - \frac{1}{2}\vec{m}_i^T\mathbf{C}^{-1}\vec{m}_i.$$

b) Ako je $\mathbf{C}_i = \sigma^2\mathbf{I}$ (disperzijske matrice su jednake i elementi vektora $\vec{x}$ su nekorelirani) funkcija odlučivanja postaje:

$$d_i(\vec{x}) = \ln p(\omega_i) - \frac{1}{2\sigma^2}||\vec{x}-\vec{m}_i||^2.$$

c) Ako pretpostavimo da uz b) vrijedi i $p(\omega_i) = 1/M$ dobivamo razvrstavanje na temelju najmanje udaljenosti:

$$d_i(\vec{x}) = \vec{x}^T\vec{m}_i - \frac{1}{2}\vec{m}_i^T\vec{m}_i.$$