Neuronske mreže: Perceptron

Prof. dr. sc. Sven Lončarić

Fakultet elektrotehnike i računarstva

https://www.fer.unizg.hr/predmet/neumre_c

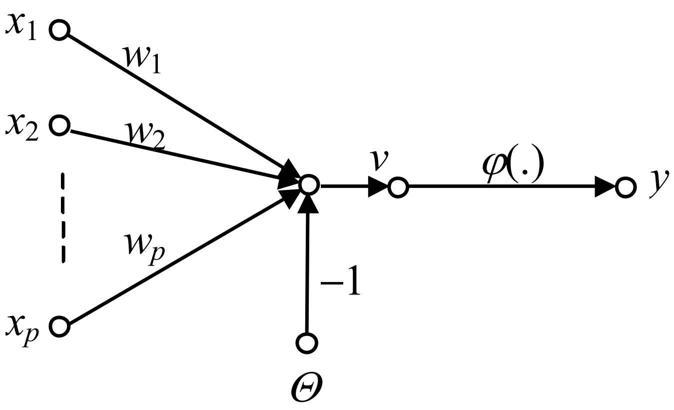
Pregled predavanja

- Uvod
- Jednoslojni perceptron
- Algoritam učenja
- Klasifikator maksimalne vjerojatnosti
- Diskusija
- Zadaci

Uvod

- Perceptron je najjednostavnija neuronska mreža za klasifikaciju uzoraka koji su linearno separabilni (koji leže na suprotnim stranama hiperravnine)
- Perceptron se sastoji od jednog neurona
- Algoritam učenja razvio je Rosenblatt, 1958
- Rosenblatt je dokazao da ako su uzorci linearno separabilni onda algoritam učenja konvergira i postavlja hiperravninu odluke točno između dvije klase (teorem konvergencije perceptrona)
- Da bi perceptron radio uzorci *moraju* biti linearno separabilni

Jednoslojni perceptron



$$v = \sum_{j=1}^{p} w_j x_j - \Theta$$
$$y = \varphi(v)$$

McCulloch-Pitts model neurona

Jednoslojni perceptron

- Perceptron služi za klasifikaciju skupa ulaznih vektora oblika $\mathbf{x} = [x_1\ x_2\ \ldots\ x_p]^T$ u jednu od dvije klase C_1 ili C_2
- Klasifikacija se odvija tako da se vektor **x** kojeg treba klasificirati dovede na ulaz perceptrona
- Ako je izlaz perceptrona:
  - $y = 1$ onda vektor **x** pripada klasi C_1
  - $y = -1$ onda vektor **x** pripada klasi C_2

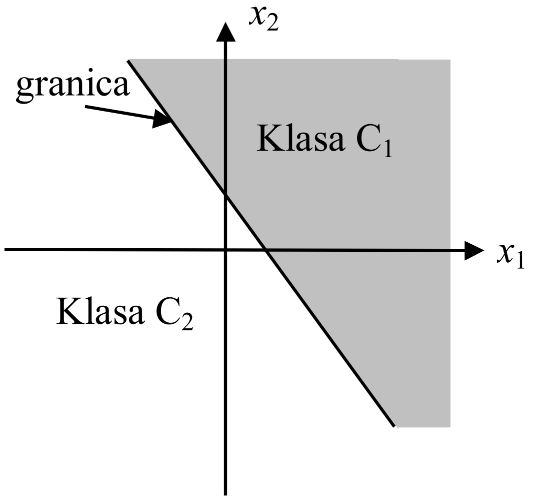
Područja klasifikacije

- Da bi se prikazala područja klasifikacije C_1 i C_2 može se promatrati iznos varijable *v* u ovisnosti o *p* ulaznih varijabli $x_1, x_2, \ldots, x_p$
- U slučaju elementarnog perceptrona postoje dva područja odvojena hiperravninom definiranom izrazom:

$$\sum_{j=1}^{p} w_j x_j - \Theta = 0$$

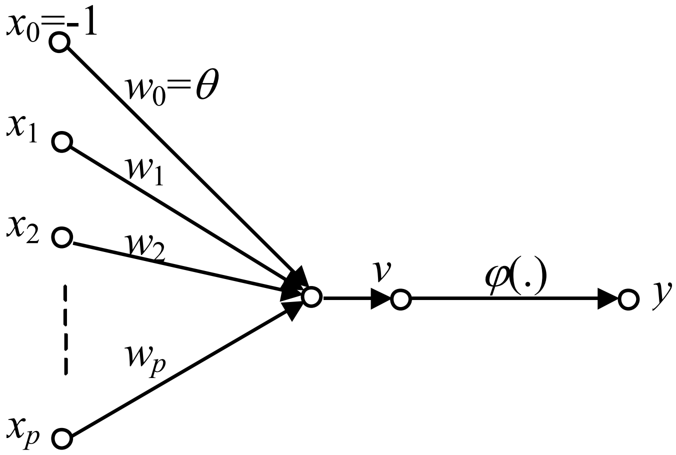
Primjer klasifikacije u dvije dimenzije

- Kod 2-D klasifikacije granica između klasa definirana je pravcem $w_1x_1+w_2x_2-\Theta = 0$
- Točka koja leži iznad pravca pripada klasi C_1
- Točka ispod pravca pripada klasi C_2



Jednoslojni perceptron

- Prag neurona možemo prikazati kao dodatni ulaz s fiksnim iznosom -1 i pripadnom težinom θ



Jednoslojni perceptron

- Definirajmo p+1 dimenzionalni ulazni vektor kao:

$$\mathbf{x}(n) = [\ -1 \ \ x_1(n) \ \ x_2(n) \ \ldots \ x_p(n)\ ]^\mathrm{T}$$

- Definirajmo p+1 dimenzionalni vektor težina kao:

$$\mathbf{w}(n) = [\theta(n) \ \ w_1(n) \ \ w_2(n) \ \ldots \ w_p(n)\ ]^\mathrm{T}$$

- Interna aktivnost neurona $v(n)$ dana je izrazom za linearnu kombinaciju ulaza i težina (skalarni produkt vektora):

$$v(n) = \mathbf{w}^T(n)\,\mathbf{x}(n)$$

- Za fiksni n jednadžba $\mathbf{w}^T(n)\,\mathbf{x}(n) = 0$ definira hiperravninu u p-dimenzionalnom prostoru koordinata x_1, x_2, …, x_p

Klasifikacija

- Ako su dvije klase uzoraka linearno separabilne onda postoji vektor težina **w** tako da vrijedi:

  $$\mathbf{w}^T\mathbf{x} >= 0$$

  za svaki vektor **x** koji pripada klasi C_1 i

  $$\mathbf{w}^T\mathbf{x} < 0$$

  za svaki vektor **x** koji pripada klasi C_2
- Problem učenja sastoji se u tome da se odredi vektor težina **w** koji će omogućiti korektnu klasifikaciju

Algoritam učenja

1. Ako je n-ti vektor $\mathbf{x}(n)$ korektno klasificiran, težina $\mathbf{w}(n)$ se ne mijenja:
   - $\mathbf{w}(n+1) = \mathbf{w}(n)$

     ako je $\mathbf{w}(n)^T \mathbf{x}(n) >= 0$ i $\mathbf{x}(n)$ pripada klasi C_1
   - $\mathbf{w}(n+1) = \mathbf{w}(n)$

     ako je $\mathbf{w}(n)^T \mathbf{x}(n) < 0$ i $\mathbf{x}(n)$ pripada klasi C_2

Algoritam učenja (nastavak)

2. Inače težinski vektor $\mathbf{w}(n)$ se mijenja na slijedeći način:

- $\mathbf{w}(n+1) = \mathbf{w}(n) - \eta(n)\,\mathbf{x}(n)$
  ako je $\mathbf{w}(n)^T\,\mathbf{x}(n) >= 0$ i $\mathbf{x}(n)$ pripada klasi C_2
- $\mathbf{w}(n+1) = \mathbf{w}(n) + \eta(n)\,\mathbf{x}(n)$
  ako je $\mathbf{w}(n)^T\,\mathbf{x}(n) < 0$ i $\mathbf{x}(n)$ pripada klasi C_1

gdje je $\eta(n)$ pozitivni parametar koji određuje brzinu učenja

Klasifikator maksimalne vjerojatnosti

- engl. maximum likelihood (ML) classifier
- Pomoću jednoslojnog perceptrona može se realizirati klasifikator koji radi na principu maksimalne vjerojatnosti (ML klasifikator)
- Problem klasifikacije uzoraka može se promatrati kao problem estimacije klase kojoj nepoznati uzorak pripada

Klasifikator maksimalne vjerojatnosti

- To je problem estimacije parametara (engl. parameter estimation)
- Parametri su veličine koje su fiksne ali nepoznate (u slučaju klasifikatora nepoznati parametar koji želimo procjeniti je indeks klase kojoj uzorak pripada)

Estimacija parametara

- Zamislimo da imamo skup uzoraka koji možemo prema klasama raspodjeliti u podskupove $X_1, X_2, \ldots, X_M$
- Neka je funkcija gustoće vjerojatnosti uzoraka $\mathbf{x}$ za pojedine klase dana izrazima $f(\mathbf{x}|\mathbf{z}_j)$ gdje je $\mathbf{z}_j$ nepoznati vektor parametara koji opisuje klasu C_j
- $f(\mathbf{x}|\mathbf{z})$ zove se vjerojatnost (engl. likelihood) od $\mathbf{z}$ s obzirom na opaženi vektor $\mathbf{x}$
- ML estimacija parametra $\mathbf{z}$ je neka vrijednost $\mathbf{z}'$ koja maksimizira vjerojatnost $f(\mathbf{x}|\mathbf{z})$

ML klasifikator

- Neka je uzorak opisan p-dimenzionalnim slučajnim vektorom $\mathbf{x}$ koji ima vektor srednje vrijednosti $\boldsymbol{\mu} = E[\mathbf{x}]$ i kovarijancijsku matricu $\mathbf{C} = E[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T]$
- Ako pretpostavimo da slučajni vektor $\mathbf{x}$ ima Gaussovu razdiobu onda je funkcija gustoće vjerojatnosti dana izrazom:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}(\det \mathbf{C})^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$

ML klasifikator

- Pretpostavimo radi ilustracije da imamo problem s dvije klase (M=2) i da je vektor uzorka **x** karakteriziran slijedećim parametrima zavisno da li pripada klasi C_1 ili C_2
- Ako uzorak **x** pripada klasi C_1:

  srednja vrijednost = $\boldsymbol{\mu}_1$ i kovarijancijska matrica = **C**
- Ako uzorak **x** pripada klasi C_2:

  srednja vrijednost = $\boldsymbol{\mu}_2$ i kovarijancijska matrica = **C**

ML klasifikator

- Problem ML estimacije parametara se sada sastoji u tome da za neki opaženi vektor **x** trebamo procjeniti da li se veća vjerojatnost pojave tog vektora dobiva za vrijednost parametra $\boldsymbol{\mu}_1$ ili za $\boldsymbol{\mu}_2$

ML klasifikator

- Za dane dvije klase možemo pisati pripadne funkcije gustoće vjerojatnosti kao:

$$f(\mathbf{x} \mid C_i) = \frac{1}{(2\pi)^{p/2} (\det \mathbf{C})^{1/2}} \exp\left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_i)^T \mathbf{C}^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) \right]$$

- Dakle treba odrediti za dani **x** koja je od vjerojatnosti $f(\mathbf{x}|C_1)$ i $f(\mathbf{x}|C_2)$ veća
- Da bi to odredili radi pojednostavljenja možemo promatrati i logaritme vjerojatnosti $\ln f(\mathbf{x}|C_1)$ i $\ln f(\mathbf{x}|C_2)$

ML klasifikator

- Logaritmi vjerojatnosti dani su izrazom gdje su samo zadnja dva pribrojnika ovisna o indeksu klase *i*:

$$\ln f(\mathbf{x}\,|\,C_i) = -\frac{p}{2}\ln(2\pi) - \frac{1}{2}\ln(\det\mathbf{C}) - \frac{1}{2}\mathbf{x}^T\mathbf{C}^{-1}x + \boldsymbol{\mu}_i{}^T\mathbf{C}^{-1}\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}_i{}^T\mathbf{C}^{-1}\boldsymbol{\mu}_i$$

- Dakle za usporedbu dvaju logaritama vjerojatnosti dovoljno je promatrati izraze:

$$l_1(\mathbf{x}) = \boldsymbol{\mu}_1{}^T\mathbf{C}^{-1}\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}_1{}^T\mathbf{C}^{-1}\boldsymbol{\mu}_1$$

$$l_2(\mathbf{x}) = \boldsymbol{\mu}_2{}^T\mathbf{C}^{-1}\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}_2{}^T\mathbf{C}^{-1}\boldsymbol{\mu}_2$$

ML klasifikator

$$l(\mathbf{x}) = l_1(\mathbf{x}) - l_2(\mathbf{x}) = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \mathbf{C}^{-1}\mathbf{x} - \frac{1}{2}\left(\boldsymbol{\mu}_1^{\ T}\mathbf{C}^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2^{\ T}\mathbf{C}^{-1}\boldsymbol{\mu}_2\right)$$

- Razlika $l(\mathbf{x})$ pokazuje koja je vjerojatnost veća:
  - ako je $l(\mathbf{x}) >= 0$ onda je $f(\mathbf{x}|C_1)$ veća vjerojatnost (klasa C_1)
  - ako je $l(\mathbf{x}) < 0$ onda je $f(\mathbf{x}|C_2)$ veća vjerojatnost (klasa C_2)
- Može se vidjeti da je veza između $l(\mathbf{x})$ i $\mathbf{x}$ linearna:

$$l(\mathbf{x}) = \mathbf{w}^T\mathbf{x} - \Theta$$

gdje je:

$$\mathbf{w} = \mathbf{C}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$$

$$\Theta = \frac{1}{2}\left(\boldsymbol{\mu}_1^{\ T}\mathbf{C}^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2^{\ T}\mathbf{C}^{-1}\boldsymbol{\mu}_2\right)$$

ML klasifikator

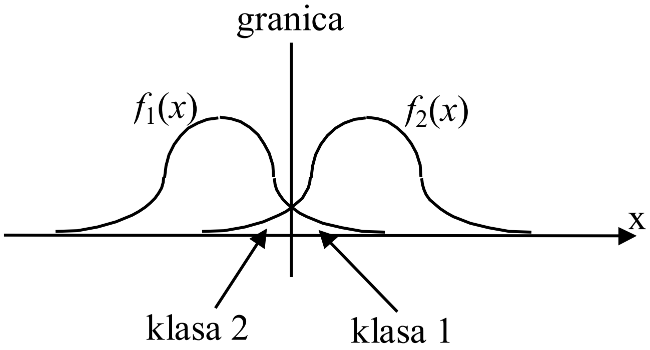
- Dakle ML klasifikator može se realizirati pomoću perceptrona koji ima vektor težina **w** i prag θ
- Interna aktivnost takvog perceptrona jednaka je:

$$l = \mathbf{wx} - \theta$$

- Klasifikacija nepoznatog uzorka **x** se obavlja na slijedeći način:
  - ako je $l > 0$ onda je $l_1 > l_2$ i znači da **x** pripada klasi C_1
  - ako je $l < 0$ onda je $l_1 < l_2$ i znači da **x** pripada klasi C_2

ML klasifikator i perceptron

- ML klasifikator i perceptron su linearni klasifikatori
- ML klasifikator je izveden uz pretpostavku da se klase preklapaju (zato se klase ne mogu točno separirati) dok perceptron radi uz pretpostavku da su klase separabilne

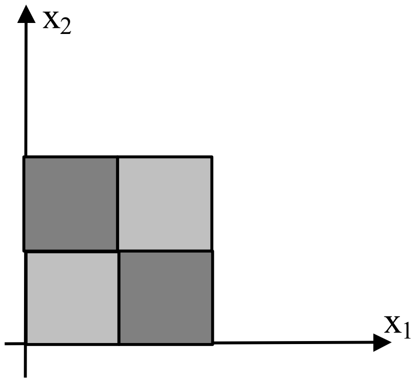


ML klasifikator i perceptron

- Perceptron ne pretpostavlja nikakve distribucije dok za ML klasifikaciju treba pretpostaviti (tj. znati) distribucijske funkcije ulaznih uzoraka
- Učenje perceptrona je adaptivno i jednostavnije za realizaciju dok je dizajn adaptivnog Gaussovog ML klasifikatora složeniji

Diskusija

- Minski je kritizirao Rosenblattov perceptron da ne može naučiti ni tako jednostavnu funkciju kao što je XOR
- To nije moguće jer dva skupa točaka na slici nisu linearno separabilni



Zadaci

- Problem 4.1.
  - Jednoslojni perceptron je linearni klasifikator. Objasniti ovu tvrdnju.
- Problem 4.2.
  - Dane su dvije jednodimenzionalne klase C1 i C2 s Gaussovim distribucijama koje imaju varijancu jednaku 1 i srednje vrijednosti μ_1=-10 i μ_2=10. Izračunati klasifikator za separaciju ovih klasa.