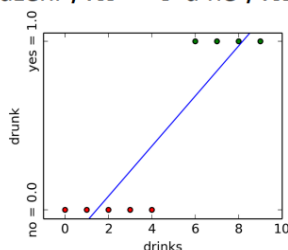


11. Binární lineární klasifikace a logistická regrese, učení modelu logistické regrese.

Binární lineární klasifikace

- V případě klasifikace požaduje výstup systému ve smyslu rozhodnutí o příslušnosti k jedné ze dvou definovaných tříd
- Přiřadíme datům z jedné třídy požadovanou výstupní hodnotu **+1** a datům z druhé třídy **0** (příp. -1)
- Stanovíme parametry lin. Regrese

Ta ovšem provádí zobrazení $f: X \rightarrow Y$ a ne $f: X \rightarrow C$



- Rozhodnutí o příslušnosti ke třídě provedeme na základě porovnání hodnoty funkce f s daným prahem
- **Za výstup z regrese tedy přidáme nelineární rozhodovací člen**

- Platí

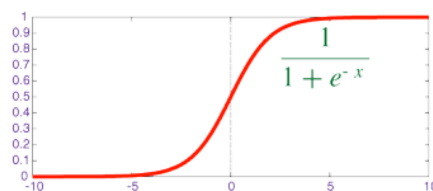
$$c_x = \begin{cases} \text{'yes'} & \text{pro } f(x) \geq 0.5 \\ \text{'no'} & \text{pro } f(x) < 0.5 \end{cases}$$

- **Nevýhoda: všechna data ovlivňují polohu separující nadroviny**
- Je třeba změnit i trénovací kritérium (cost funkci)
 - Nechceme aproximovat body ale separovat třídy !!

Logistická regrese

Funkce sigmoida:

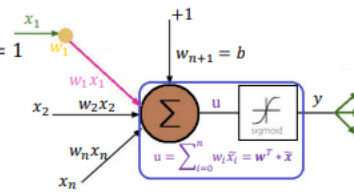
- Funkce definovaná jako:
 - $\text{sigmoid}(x) = \sigma(x) = \frac{e^x}{1+e^x} = \frac{1}{1+e^{-x}}$
 - Má zajímavé vlastnosti:
 - definiční obor $(-\infty, +\infty)$
 - obor hodnot $(0,1)$
- ⇒ „přiřazuje reálným číslům pravděpodobnost“



$$\sigma(x) > 0.5 \text{ pro } x > 0 \quad \sigma(x) < 0.5 \text{ pro } x < 0$$

Binární lineární klasifikátor se sigmoidou #1

- Vstupem je vektor $\mathbf{x} = [x_1; x_2; \dots; x_n]$
 - Například hodnoty pixelů testovacího obrázku
- Každá vstupní hodnota x_i se vynásobí vahou w_i , výsledek se sečte a přičte se k němu váha w_{n+1}
 - To lze zapsat jako $\sum_{i=1}^{n+1} w_i \tilde{x}_i = \mathbf{w}^T \tilde{\mathbf{x}} = u$
 - Kde $\tilde{\mathbf{x}}$ je původní vektor \mathbf{x} rozšířený o hodnotu $x_{n+1} = 1$
- Hodnota tohoto součinu říká, zda bod \mathbf{x} leží v prostoru o n dimenzích a souřadných osách x_1, x_2, \dots, x_n nad, pod nebo přímo na podploše o dimenzi $n - 1$, která je definována rovnicí $w_1 x_1 + \dots + w_n x_n + w_{n+1} = 0$
- Hodnota součinu u se dosadí do sigmoidy
 - "Převéde se na pravděpodobnost v rozmezí (0,1)"



Binární lineární klasifikátor se sigmoidou #2

- Pokud je hodnota na výstupu sigmoidy ≥ 0.5 , přiřadíme \mathbf{x} ke třídě 1
- Pokud je hodnota na výstupu sigmoidy < 0.5 , přiřadíme \mathbf{x} ke třídě 0
- Popsaný klasifikátor je binární a pravděpodobnostní**
 - Třída 1 má pro \mathbf{x} pravděpodobnost $\sigma(u) = \sigma(\mathbf{w}^T \tilde{\mathbf{x}})$
 - Třída 0 má pro \mathbf{x} pravděpodobnost $1 - \sigma(u) = 1 - \sigma(\mathbf{w}^T \tilde{\mathbf{x}})$
- Umí klasifikovat jen třídy separovatelné lineárně (přímku)**
 - $\mathbf{w}^T \tilde{\mathbf{x}} \geq 0 \Rightarrow \sigma(\mathbf{w}^T \tilde{\mathbf{x}}) \geq 0.5 \Rightarrow$ bod leží v jedné části příznakového prostoru
 - $\mathbf{w}^T \tilde{\mathbf{x}} < 0 \Rightarrow \sigma(\mathbf{w}^T \tilde{\mathbf{x}}) < 0.5 \Rightarrow$ bod leží ve druhé části příznakového prostoru
- Klasifikátor je parametrický s vektorem parametrů \mathbf{w}**
 - Během trénování se používá jiné kritérium než pro učení obyčejné regrese !!

Učení modelu logistické regrese

- Pravděpodobnost náhodného jevu lze určit kombinatorickým rozbořem nebo z dat pomocí relativní četnosti
- Náhodná veličina je výsledek náhodného pokusu vyjádřený číslem
- Učení Logistické regrese pomocí numerické řešení metody **MLE** (Maximum Likelihood Estimation) pro logistickou regresi pomocí SGD:

- Co je to Metoda maximální věrohodnosti?
 - postup, jak z trénovacích dat odhadnout (určit) parametry modelu
- Cílem je najít parametry maximalizující věrohodnost dat
 - tj. "pravděpodobnost" toho, že daný model tato data vygeneroval
- Výsledný odhad má celou řadu pozitivních vlastností:
 - je asymptoticky eficientní (pro počet pozorování $n \rightarrow \infty$)
 - tj. odhadujeme neznámý parametr nejlepším možným způsobem.
- Existují i jiné metody odhadu?
 - ano, například Metoda nejmenších čtverců = Least Squares Estimation (LSE)

Obecný postup při aplikaci metody MLE

- 1) Vyjádří se věrohodnost dat = sdružené rozdělení pravděpodobnosti celého souboru dat
 - pro soubor stejně rozdělených, nezávislých diskrétních náhodných veličin platí, že jejich sdruženou distribuci lze rozdělit na součin jednotlivých rozdělení

$$p(X_1, X_2, \dots, X_N | \theta) = p(X_1 | \theta) p(X_2 | \theta) \dots p(X_N | \theta) = \prod_{i=1}^N p(X_i | \theta)$$

- 2) Vypočte se logaritmus této věrohodnosti

$$\ln p(X_{1:N}) = \ln \prod_{i=1}^N p(X_i | \theta) = \sum_{i=1}^N \ln p(X_i | \theta)$$

- 3) Hledá se takový odhad $\hat{\theta}$, který tuto věrohodnost maximalizuje

$$\hat{\theta}_{MLE} = \underset{\theta \in \Theta}{\operatorname{argmax}} \sum_{i=1}^N \ln p(X_i | \theta)$$

$\hat{\theta}_{MLE}$ se najde tak, že vztah pro věrohodnost se zderivuje a položí roven nule

Podrobnější postup:

- 1) Vyjádří se věrohodnost dat = sdružené rozdělení pravděpod. celého souboru dat

$$\prod_{i=1}^N p(X_i | \theta) = \prod_{i=1}^N \theta^{X_i} (1 - \theta)^{1-X_i}$$

X ale nabývá pouze hodnot {0,1}

$$\prod_{i=1}^N \theta^{X_i} (1 - \theta)^{1-X_i} = \theta^k (1 - \theta)^{N-k}$$

kde k je počet hodnot {1} v souboru dat

- 2) Vypočte se logaritmus této věrohodnosti

$$\ln(\theta^k (1 - \theta)^{N-k}) = k \ln \theta + (N - k) \ln(1 - \theta)$$

- 3) Vypočte se derivace a položí se rovna nule

$$\frac{d}{d\theta} (k \ln \theta + (N - k) \ln(1 - \theta)) = \frac{k}{\theta} + \frac{(N - k)(-1)}{1 - \theta} = 0$$

$$k(1 - \theta) = (N - k)\theta$$

$$k - k\theta = N\theta - k\theta$$

$$\theta = \frac{k}{N}$$

Výsledná kritériální funkce:

Vlastnosti kritériální funkce

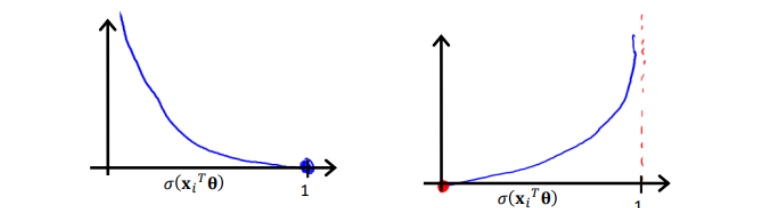
- Maximalizovat cost funkci znamená minimalizovat fci

$$J(\theta) = - \sum_{i=1}^N \{y_i \ln(\sigma(\mathbf{x}_i^T \theta)) + (1 - y_i) \ln(1 - \sigma(\mathbf{x}_i^T \theta))\}$$

- Zároveň platí, že y nabývá pouze hodnot 0 nebo 1

Pro $y = 1$ minimalizujeme
 $-\ln(\sigma(\mathbf{x}_i^T \theta))$, druhý člen = 0

Pro $y = 0$ minimalizujeme
 $-\ln(1 - \sigma(\mathbf{x}_i^T \theta))$, první člen = 0



- Řešení metodou SGD:

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \alpha \sum_{i=1}^N \mathbf{x}_i (\sigma(\mathbf{x}_i^T \boldsymbol{\theta}_t) - y_i)$$

kde α je velikost kroku (learning rate)

- Trénování se zastavuje pokud
 - Se dostatečně nemění hodnoty $\boldsymbol{\theta}$ mezi dvěma iteracemi
 - Dochází ke zhoršení přesnosti rozpoznávání na evaluační sadě