RPZ Test Leden 2022

Jméno:

1. (6b) Bayesovská klasifikace.

   Předpokládejme, že máme dvě třídy hub, jedlé (E) a toxické (T). Automatický systém navržený tak, aby pomohl s rozpoznáním těchto tříd, poskytuje výstup, který je 0 (velmi nízké riziko otravy), 1 (střední riziko otravy) nebo 2 (vysoké riziko otravy).
   Apriorní pravděpodobnosti pro uvedené dvě třídy jsou

   $$p(E) = 0.99, p(T) = 0.01\,.$$

   Podmíněné pravděpodobnosti jsou

   $$\begin{array}{lll} p(0|E) = 0.89, & p(1|E) = 0.10, & p(2|E) = 0.01 \\ p(0|T) = 0.01, & p(1|T) = 0.20, & p(2|T) = 0.79. \end{array}$$

   (a) (2b) Vypočítejte podmíněnou pravděpodobnost $p(T|1)$, že houba je toxická za předpokladu, že systém hlásí střední riziko.

   (b) (2b) Cílem je učinit binární rozhodnutí Y (*yes*, houba je akceptována) nebo N (*no*, houba se vyhodí). Penalizační funkce je:

   $$\begin{array}{ll} W(E,Y) = 0, & W(E,N) = 1, \\ W(T,Y) = 1000, & W(T,N) = 0. \end{array}$$

   Určete bayesovskou optimální strategii, tedy optimální rozhodnutí pro všechny výstupy systému (0, 1, 2).

   (c) (2b) Popište krátce formálně (pomocí vzorců) jak se optimální bayesovská strategie spočítá.

2. (6b) Odhad parametrů.
   Kostkou se hází opakovaně tolikrát, dokud se na horní straně neobjeví „šestka“. Pravděpodobnost, že se to stane při n-tém hodu, je

   $$p(n) = (1-q)^{n-1}q\,, \tag{2}$$

   kde q je pravděpodonost, že na kostce padne „šestka“.

   (a) (1b) Zdůvodněte správnost vzorce pro $p(n)$.

   (b) (3b) Jaký je maximálně vědohodný odhad pro q pokud šestka padla v k-tém hodu?

   (c) (2b) Předpokládejme, že kostkou hodíme maximálně n_{max}-krát. Jaký je maximálně věrohodný odhad pro q, když šestka ani v posledním hodu nepadne?

3. (4b) Neuronové sítě.

   (a) (1b) Vysvětlete, co je vanishing gradient a jak se tento problém řeší.

   (b) (1b) Co je max-pooling? Proč se používá?

   (c) (1b) Co je softmax?

   (d) (1b) Co je metoda stochastic gradient descent?

4. (4b) Vysvětlete, co je K-means++. Co přináší použití K-means++ ve srovnání se standardním K-means?

$$p(T|1) = \frac{p(1|T)\,p(T)}{p(1)} = \frac{0.2 \cdot 0.01}{0.101} = 0.0198$$

sdružené:

	0	1	2	Σ
E	0.8811	0.099	0.0099	0.99
T	0.0001	0.002	0.0079	0.01
Σ	0.8812	0.101	0.0178	1
	p(0)	p(1)	p(2)	

$W(k,d)$

	$k=E$	$k=T$
$d=Y$	0	1000
$d=N$	1	0

pro 0: Y $p(T,0)\cdot 1000 = 0.1$
N $p(E,0)\cdot 1 = 0.8811$

pro 1: Y $p(T,1)\cdot 1000 = 2$
N $p(E,1)\cdot 1 = 0.099$

pro 2: Y $p(T,2)\cdot 1000 = 7.9$
N $p(E,2)\cdot 1 = 0.0099$

$$R(q) = \sum_x \sum_h p(x,h)\,W(h,q(x))$$

$$q(k) = \begin{cases} Y & k=0 \\ N & k=1 \\ N & k=2 \end{cases}$$

optimální bayesovská strategie minimalizuje riziko, tedy hledá $q^*(x)$ takové:

$$q^*(x) = \underset{q}{\operatorname{argmin}} \sum_{x \in X} \sum_{h \in H} p(x,h)\,W(h,q(x))$$

(2)

$$p(n) = (1-q)^{n-1} q$$

(a) pst. že prvně šestka: q

pst. že nejprve šestka $1-q$

pst. že nejprve šestka po x hodech $(1-q)^x$

pst. že nejprve šestka po x hodech a pak první šestka:

$(1-q)^x q$

potom celkový počet hodů je n, kde $x = n-1$

$(1-q)^{n-1} q$ $\left(\prod_{i=0}^{n-1} (1-q) \right) \cdot q$

(b) $L(q) = \prod_{i=0}^{k} (1-q)^{k-1} q = (1-q)^{k^2-k} q^k$

$$\ell(q) = \log (1-q)^{k^2-k} + \log q^k$$

$$= (k^2-k)\log(1-q) + k \log q$$

$$\frac{\partial \ell}{\partial q} = -\frac{k^2-k}{1-q} + \frac{k}{q} = 0 \quad / \cdot (1-q)q$$

$$-q(k^2-k) + (1-q)k = 0$$

$$-qk^2 + qk + k - qk = 0$$

$$qk^2 = k$$

$$q = \frac{1}{k}$$

$(1-q)^{n-1} q$

$(n-1)\log(1-q) + \log q$

$\frac{\partial \ell}{\partial q} = -\frac{n-1}{1-q} + \frac{1}{q} = 0$

$-kq + q + 1 - q = 0$

$-kq + 1 = 0$

$q = \frac{1}{k}$

(c) $L(q) = \prod_{h_{max}} (1-q) = (1-q)^{h_{max}}$ $q \in [0,1]$ $(1-q)^{h_{max}}$

$l(q) = \log(1-q)^{h_{max}} = h_{max} \cdot \log(1-q)$ $h_{max} \log(1-q)$

$\frac{\partial l}{\partial q} = -\frac{h_{max}}{1-q} = 0$ $\frac{\partial l}{\partial q} = -\frac{h_{max}}{1-q} = 0$

↓ max $v q = 0$

↪ min v $q = 0$

(3) softmax $p(h|x) = \frac{e^{w_h \cdot x}}{\sum_{i=1}^{h} e^{w_i x}}$

aktivační funkce, největšího má hrana NN

zvýrazní největší prvek oproti ostatním

max pooling → jeden ze skrytých vrstev, vezme výstupy

maximální prvek a ten propaguje dále

může snížit dimenzi, umožňuje lehčí výpočty

vanishing gradient — velmi malé změny při výpočtu gradientu,

síť se poté přestává učit. často při použití sigmoidy

(velká změna na vstupu zapříčiní malou na výstupu)

řešení (lepší inicializace vah a lepší aktivační fce (eg. parametric ReLU)

stochastic gradient descent — iterativní optimalizační metoda

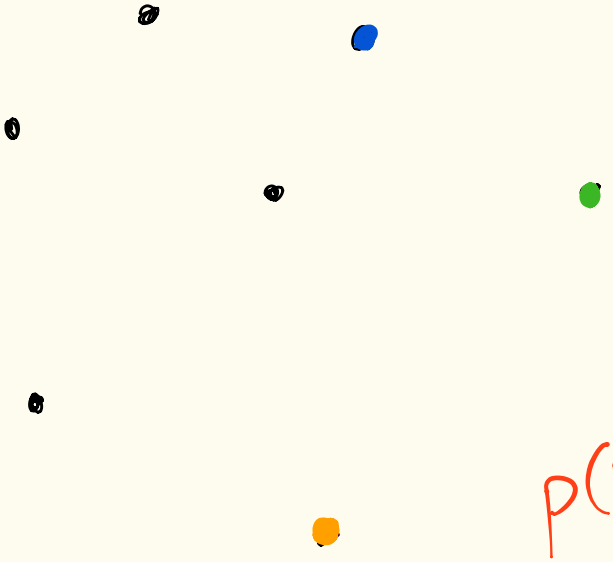
na rozdíl od gradient descent nepočítá celý dataset ale

vybere subset

je rychlejší

$$d_l = \min_{c \in C} \|x_l - c\|_2$$

$$p(l) = \frac{d_l^2}{\sum_{l \in L} d_l^2}$$



$$p(s) = \frac{d_s^2}{\sum_{1..5} d_i^2}$$