HW1 Tao Ni

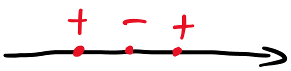
2019年2月21日 星期四 下午5:10

1\. (a) 

The threshold classifier: $f(x) = \begin{cases} +1 & x > c \\ -1 & x < c \end{cases}$

cannot shatter the configuration shown above.

$\therefore \quad VC(c) = 1$

(b) 

In this configuration, the interval classifier always has to contain the negative point if he tries to contain all the positive points.

Therefore, $VC([a,b]) = 2$.

2\. (a) $l(\theta | x) = f(x_1, x_2, \ldots, x_n | \theta) = \prod_{i=1}^{n} \frac{1}{\theta} e^{-\frac{x_i}{\theta}}$

$$\mathcal{L}(\theta | x) = -\sum_{i=1}^{n} \lg\theta - \sum_{i=1}^{n} \frac{x_i}{\theta} = -n\lg\theta - \sum_{i=1}^{n} \frac{x_i}{\theta}$$

$$\underset{\theta}{\operatorname{argmax}} \mathcal{L}(\theta | x) = \frac{\partial \mathcal{L}(\theta | x)}{\partial \theta} = 0$$

$$\Longrightarrow \quad -\frac{n}{\theta} + \sum_{i=1}^{n} x_i \cdot \frac{1}{\theta^2} = 0 \Longrightarrow \theta = \frac{1}{n} \sum_{i=1}^{n} x_i$$

(b) $l(\theta | x) = f(x_1, x_2, \ldots, x_n | \theta) = \prod_{i=1}^{n} 2\theta \cdot x_i^{2\theta - 1}$

$$\mathcal{L}(\theta | x) = n\lg 2\theta + (2\theta - 1) \sum_{i=1}^{n} \lg x_i$$

$$\underset{\theta}{\operatorname{argmax}} \mathcal{L}(\theta | x) = \frac{\partial \mathcal{L}(\theta | x)}{\partial \theta} = 0$$

$$\Longrightarrow \frac{2n}{2\theta} + 2\sum_{i=1}^{n} \lg x_i = 0 \Longrightarrow \theta = \frac{-n}{2\sum_{i=1}^{n} \lg x_i}$$

(c) $l(\theta|x) = \prod_{i=1}^{n} \frac{1}{2\theta} = \frac{1}{(2\theta)^n}$

to maximize $l(\theta|x)$ equals to minimize θ.

$\because\ x \le 2\theta \quad \therefore\ \theta = \max\{\frac{x_1}{2}, \frac{x_2}{2}, \cdots, \frac{x_n}{2}\}$

3. (a) $P(C_1|x=0) \propto P(x=0|C_1) \cdot P(C_1) = p_1 \cdot P(C_1)$

$P(C_2|x=0) \propto P(x=0|C_2) \cdot P(C_2) = p_2 \cdot P(C_2)$

if $p_1 \cdot P(C_1) > p_2 \cdot P(C_2)$ then we classify $x=0$ as C_1

else we classify $x=0$ as C_2.

$P(C_1|x=1) \propto P(x=1|C_1) \cdot P(C_1) = (1-p_1) \cdot P(C_1)$

$P(C_2|x=1) \propto P(x=1|C_2) \cdot P(C_2) = (1-p_2) \cdot P(C_2)$

if $(1-p_1) \cdot P(C_1) > (1-p_2) \cdot P(C_2)$ then classify $x=1$ as C_1.

else classify $x=1$ as C_2.

(b) $P(C_1|\mathbf{x}) \propto P(\mathbf{x}|C_1) \cdot P(C_1) = \prod_{i=1}^{d} P_{1i}^{1-x_i} \cdot (1-P_{1i})^{x_i} \cdot P(C_1)$ ①

$P(C_2|\mathbf{x}) \propto P(\mathbf{x}|C_2) \cdot P(C_2)$

② $= P(C_2) \cdot \prod_{i=1}^{d} P_{2i}^{1-x_i} \cdot (1-P_{2i})^{x_i}$

if ① > ②, then classify $\mathbf{x}$ as C_1

else classify x as C_2.

(c) for sample $(0,0)$, $\boxed{P(C_1)=0.2}$

$$P(C_1|(0,0))=\frac{P((0,0)|C_1)P(C_1)}{P((0,0))}=\frac{0.2\times p_{11}\times p_{12}}{0.2\times p_{11}\times p_{12}+0.8\times p_{21}\times p_{22}}=0.027$$

$$P(C_2|(0,0))=1-P(C_1|(0,0))=0.973$$

$\boxed{P(C_1)=0.6}$:

$$P(C_1|(0,0))=\frac{0.6\times0.6\times0.1}{0.6\times0.6\times0.1+0.4\times0.6\times0.9}=0.143$$

$$P(C_2|(0,0))=1-0.143=0.857$$

$\boxed{P(C_1)=0.8}$

$$P(C_1|(0,0))=\frac{0.8\times0.6\times0.1}{0.8\times0.6\times0.1+0.2\times0.6\times0.9}=0.308$$

$$P(C_2|(0,0))=1-0.308=0.692$$

for others samples:

posterior	(0,1)	(1,0)	(1,1)	
$P(C_1)=0.2$	0.692	0.027	0.693	C_1
	0.308	0.973	0.307	C_2
$P(C_1)=0.6$	0.931	0.143	0.931	C_1
	0.069	0.857	0.069	C_2
$P(C_1)=0.8$	0.973	0.307	0.973	C_1
	0.027	0.693	0.027	C_2

---

4. validation_err_rate:

σ	0.00001	0.0001	0.001	0.01	0.1	1	2	3
err	0.54	0.54	0.54	0.54	0.51	0.515	0.455	0.46

σ	4	5	6
err	0.46	0.46	0.46

test_err_rate: 44.5%