2023feb17

Friday, February 17, 2023 12:58

1

		A	
		0	1
B	0	0.42	0.18
	1	0.28	0.12

$P(B)$

	~~b~~	$P(B)$
0	$0.42+0.18=$	0.6
1	$0.28+0.12$	$=0.4$

$P(A=1) = P(A=1, B=0) + P(A=1, B=1)$
$= 0.18 + 0.12$
$= 0.3$

$P(A=0)$
$= 0.7$

	0	1
$P(A)=$	0.7	0.3

$P(A=0) = 0.42 + 0.28$
$P(A=1) = 0.18 + 0.12$

b Are they independent?

Does $P(A,B) = P(A)\,P(B)$?

$P(A=0, B=1) = P(A=0)\,P(B=1)$?

$0.28 = 0.7 \times 0.4$ ✓ Yes

Yes, independent b/c

$P(A=0, B=1) = P(A=0)\,P(B=1)$

and b/c they're binary, this implies that

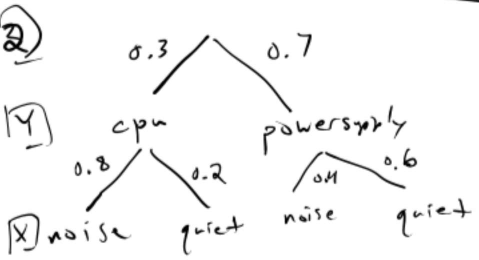
$P(A=a, B=b) = P(A=a)\,P(B=b)$ FOR ALL a, b

$P(A=1, B=1) = P(B=1) - P(A=0, B=1)$

$P(A=1) = 1 - P(A=0)$

$P(A=1, B=1) = 0.12 \stackrel{?}{=} P(A=1)\,P(B=1) = 0.3 \times 0.4$ ✓

2



		Y	
		cpu	pow
X	noise	$(0.3)(0.8)$	$(0.7)(0.4)$
	quiet	$(0.3)(0.2)$	$(0.7)(0.6)$

$$\boxed{f(X) = \operatorname{argmax}_Y P(Y \mid X) = \operatorname{argmax}_Y P(Y, X) = \begin{cases} \text{cpu} & \text{never} \\ \text{powerSupply} & \text{if } X = \text{quiet or } X = \text{noisy} \end{cases}}$$

b Bayes error rate

$= P(f(x) \neq Y)$ for MAP classifier

$= P(Y = cpu) = 0.3$

c $P(FA) = P(f(x) = cpu \mid Y = powersupply)$

$= 0$

$P(MD) = P(f(x) = powersupply \mid Y = cpu)$

$= 1$

4) FAIRNESS

$P(Y=1 \mid A=1) = \frac{2}{3}$

GIVEN $A=1$ NOT TOLD: $A=0$

Y	$P(\hat{Y}=0 \mid Y, A=1)$	$P(\hat{Y}=1 \mid Y, A=1)$
0	0.8	0.2
1	0.4	0.6

$$P(Y=1 \mid \hat{Y}=1, A=1) = \frac{P(Y=1, \hat{Y}=1 \mid A=1)}{P(\hat{Y}=1 \mid A=1)}$$

$$= \frac{P(Y=1 \mid A=1)\, P(\hat{Y}=1 \mid Y=1, A=1)}{P(Y=1 \mid A=1)\, P(\hat{Y}=1 \mid Y=1, A=1) + P(Y=0 \mid A=1)\, P(\hat{Y}=1 \mid Y=0, A=1)}$$

$$= \frac{\left(\frac{2}{3}\right)(0.6)}{\left(\frac{2}{3}\right)(0.6) + \left(\frac{1}{3}\right)(0.2)}$$

Demo. Parity $P(\hat{Y}=1 \mid A=1) = P(\hat{Y}=1 \mid A=0)$

if Y may not be fair

Equal odds $P(\hat{Y}=1 \mid Y=1, A=1) = P(\hat{Y}=1 \mid Y=1, A=0)$

if Y fair

6) Linear Regression

$$\mathcal{L} = (w @ x - y)^2$$

$$= (w_1 x_1 + w_2 x_2 + b - y)^2$$

$$\frac{\partial \mathcal{L}}{\partial w_2} = 2(w_1 x_1 + w_2 x_2 + b - y)\, \frac{\partial (w_1 x_1 + w_2 x_2 + b - y)}{\partial w_2}$$

$$= 2(w_1 x_1 + w_2 x_2 + b - y)\, x_2$$

$$= 2 \epsilon x$$

where $\epsilon = w @ x - y$

7) Linear Classifier

$$f_k = \frac{\exp(\xi_k)}{\sum_j \exp(\xi_j)}$$

$$\frac{\partial f_5}{\partial \xi_3} = \frac{1}{DEN} \cdot \frac{\partial NUM}{\partial \xi_3} - \frac{NUM}{DEN^2} \cdot \frac{\partial DEN}{\partial \xi_3}$$

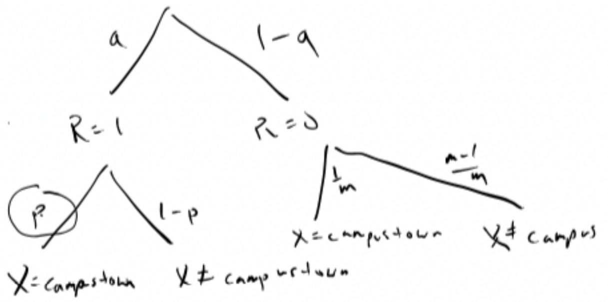
$$= \frac{1}{\sum_j \exp(\xi_j)} \cdot \cancelto{0}{\frac{\partial \exp(\xi_5)}{\partial \xi_3}} - \frac{\exp(\xi_5)}{\left(\sum_j \exp(\xi_j)\right)^2} \frac{\partial DEN}{\partial \xi_3}$$

$$= 0 - \frac{\exp(\xi_5)}{\left(\sum_j \exp(\xi_j)\right)^2} \frac{\partial \left(\sum_j \exp(\xi_j)\right)}{\partial \xi_3}$$

$$\frac{\partial e^{\xi_3}}{\partial \xi_3} = e^{\xi_3}$$

$$\frac{\partial f_5}{\partial \xi_3} = -\frac{\exp(\xi_5)\exp(\xi_3)}{\left(\sum_j \exp(\xi_j)\right)^2} = -f_5 f_3$$

10) PRIVACY



$$P(X = \text{compustown}) = b$$

$$= P(X = \text{comp}, R = 1) + P(X = \text{comp}, R = 0)$$

$$b = ap + (1-a)\frac{1}{m}$$

$$p = \frac{b - (1-a)/m}{a}$$

---

9 OPTIMIZATION

$w = [w_0, \dots, w_{m-1}]$ $\qquad w_i \in \{0, \dots, n-1\}$

1a Exhaustive search

m coefficients, each has n values

$\Rightarrow n^m$ possible w vectors

$\Rightarrow \underline{O\{n^m\}}$

1b Coordinate search w/ random restarts

For each restart:

generate a random starting w

for each iteration:

for each coordinate $0 \le i \le m-1$:

- find best possible value, $\hat{w}_i$, of w_i, under condition that w_j fixed. Record $\mathcal{L}_i = \mathcal{L}([w_0, \dots, w_{i-1}, \hat{w}_i, w_{i+1}, \dots, w_{m-1}])$

Find $i^* = \arg\min \mathcal{L}_i$

$w \leftarrow [w_0, \dots, w_{i^*-1}, \hat{w}_{i^*}, w_{i^*+1}, \dots, w_{m-1}]$

Restarts: p

Iterations: q

Coordinates: m

Values: n

Total complexity: $pqmn$ $\qquad O\{pqmn\}$

Question: How large does p have to be to make

$$O\{pqmn\} = O\{n^m\}$$

$$\boxed{p = \frac{n^m}{qmn} = \frac{n^{m-1}}{qm}}$$