Homework #4

Srujith Poondla
spoondla.

1) Given training set

	A	B	C	Class
Instance 1	F	T	120	Benign
Instance 2	T	F	1090	Benign
Instance 3	T	T	245	Malignant
Instance 4	F	F	589	Malignant
Instance 5	T	T	877	Malignant.

Information gained by knowing whether or not the value of feature C is less than 475 is :-

Entropy of class : $H(\text{class}) = -\frac{2}{5}\log(2/5) - \frac{3}{5}\log(3/5)$

$= 0.9709$

$H(\text{class}) \simeq 0.9710.$

$$\boxed{I(\text{class}; C) = H(\text{Class}) - H(\text{class}/C)}$$

$$H(\text{class}/c) = (2/5)\, H(\text{class}/c \le 375) + (3/5)\, H(\text{class}/c > 475)$$

$$= (2/5)\left(-\tfrac{1}{2}\log\tfrac{1}{2} - \tfrac{1}{2}\log(\tfrac{1}{2})\right) + (3/5)\left(-\tfrac{1}{3}\log\tfrac{1}{3} - \tfrac{2}{3}\log\tfrac{2}{3}\right.$$

$$= (2/5) + (3/5)(0.9182)$$

$$H(\text{class}/c) = 0.9509$$

(a) $I(\text{class}; c) = H(\text{class}) - H(\text{class}/c)$

$= 0.97095 - 0.95097$

$= 0.9710 - 0.9510$

$= 0.02$

$\therefore$ Information gained $= 0.02$.

(b) Information gained about the class whether or not the value of features A & B are different.

from the previous question we know

$H(\text{class}) = 0.9710$.

let's assume X – A & B same
Y – A & B different.

$$H(\text{class}/xy) = (2/5)(H(\text{class}/y)) + (3/5)(H(\text{class}/x))$$

$$= 2/5(-0-0) + (3/5)(0)$$

$$= 0.$$

$\therefore$ Information gain $= I(\text{class}; XY) = H(\text{class}) - H(\text{class}/xy)$

$= 0.9710 - 0$

$\therefore$ Information gain $= 0.9710$

Scanned by CamScanner

②

	feature 1	feature 2	class
Instance 1	2	3	Positive
Instance 2	4	4	Positive
Instance 3	4	5	Negative
Instance 4	6	3	Positive
Instance 5	8	3	Negative
Instance 6.	8	4	Negative

KNN with K=1

K=1 :-

			classification
Inst: 1	class = +ve	closest instance is 2 Manhattan distance 3	correct.
Inst: 2	class = +ve	closest instance is 3 Manhattan distance 1	correct.
Inst: 3	class = -ve	closest instance is 2 Manhatan distance 1	incorrect.
Inst: 4	class = +ve	closest instance is 5 manhattan distance 2	incorrect.
Inst: 5	class = -ve	closest instance is 6 Manhattan distance 1	correct.
Inst: 6	class = -ve	closest instance is 5 Manhattan distance 1	correct.

KNN with K=2

Instance	class	closest Instance	Manhattan distance	classification.
1	+ve	2	3	Correct
		3	4	incorrect.
2	+ve	3	1	incorrect
		1	3	correct
3	−ve	2	1	In correct
		1	4	incorrect
4	+ve	5	2	In correct
		2	3	Correct.
5	−ve	6	1	correct
		4	2	incorrect.
6	−ve	5	1	correct
		4	3	incorrect.

No. of correctly classified = 3

KNN with k = 3.

Instance	class	closest Instance	Manhattan distance	Classification.
1	+ve	2	3	correct
		3	4	incorrect
		4	4	correct
2	+ve	3	1	incorrect
		1	3	correct
		4	3	correct.
3	−ve	2	1	incorrect
		1	4	incorrect
		4	4	incorrect.
4	+ve	5	2	incorrect
		2	3	correct
		6	3	incorrect.
5	−ve	6	1	correct
		4	2	incorrect
		2	5	incorrect.
6	−ve.	5	1	correct
		4	3	incorrect
		2	4	incorrect.

Total correct classified instancy = 2

K = 1 & K = 2 both gave same results.

(3)

X	Y	Z	count
T	T	T	36
T	T	F	4
T	F	T	2
T	F	F	8
F	T	T	9
F	T	F	1
F	F	T	8
F	F	F	32

Mutual information between Z & X i.e $I(X,Z)$

3.a.1) $P(X=T, Z=T) \log \frac{P(X=T, Z=T)}{P(X=T)P(Z=T)} = 0.38 \log \frac{0.38}{0.5 \times 0.55}$

$= 0.177$

$P(X=T, Z=F) \log_2 \frac{P(X=T, Z=F)}{P(X=T)\,P(Z=F)} = 0.12 \log \frac{0.12}{0.5 \times 0.45} =$

Scanned by CamScanner

$$P(x=F, z=T) \log_2 \frac{P(x=F, z=T)}{P(x=F)\cdot P(z=T)} = 0.17 \log \frac{0.17}{0.5 \times 0.55}$$

$$P(x=F, z=F) \log_2 \frac{P(x=F, z=F)}{P(x=F)\, P(z=F)} = 0.33 \log \frac{0.33}{0.5 \times 0.45}$$

$$I(x, z) = \sum_{x,z} P(x, z) \log_2 \frac{P(x,z)}{P(x)\, P(z)}$$

$$I(x, z) = 0.133$$

Q) 2)

Mutual Information between Z & Y

$$P(Y=T, Z=T) \log_2 \frac{P(Y=T, Z=T)}{P(Y=T)\, P(Z=T)} = 0.45 \log_2 \frac{0.45}{0.5 \times 0.55}$$

$$P(Y=T, Z=F) \log_2 \frac{P(Y=T, Z=F)}{P(Y=T)\, P(Z=F)} = 0.05 \log_2 \frac{0.05}{0.5 \times 0.45}$$

$$P(Y=F, Z=T) \log_2 \frac{P(Y=F, Z=T)}{P(Y=F)\, P(Z=F)} = 0.1 \log_2 \frac{0.1}{0.6 \times 0.55}$$

$$P(Y=F, Z=F) \log_2 \frac{P(Y=F, Z=F)}{P(Y=F)\, P(Z=F)} = 0.4 \log_2 \frac{0.4}{0.5 \times 0.45}$$

$$I(Y, Z) = 0.398$$

3.b)

Y should be selected as the parent for Z as it has better mutual information between Z & Y than Z & X.

9. P(X)

T	F
0.5	0.5

P(Y/X)

Y/X	T	F
T	0.8	0.2
F	0.2	0.8

P(Z/Y)

Z/Y	T	F
T	0.9	0.1
F	0.2	0.8

4

X coordinate	Y coordinate
12	4
3	18
6	11
5	5

a) polynomial Kernel of degree 2.

$x^{(1)} = (12, 4)$
$x^{(2)} = (3, 18)$
$x^{(3)} = (6, 11)$
$x^{(4)} = (5, 5)$

$$\begin{bmatrix} K(x^{(1)}, x^{(1)}) & K(x^{(1)}, x^{(2)}) & K(x^{(1)}, x^{(3)}) & K(x^{(1)}, x^{(4)}) \\ K(x^{(2)}, x^{(1)}) & K(x^{(2)}, x^{(2)}) & K(x^{(2)}, x^{(3)}) & K(x^{(2)}, x^{(4)}) \\ K(x^{(3)}, x^{(1)}) & K(x^{(3)}, x^{(2)}) & K(x^{(3)}, x^{(3)}) & K(x^{(3)}, x^{(4)}) \\ K(x^{(4)}, x^{(1)}) & K(x^{(4)}, x^{(2)}) & K(x^{(4)}, x^{(3)}) & K(x^{(4)}, x^{(4)}) \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} (160)^2 & (36+72)^2 & (116)^2 & (80)^2 \\ (108)^2 & (333)^2 & (216)^2 & (105)^2 \\ (116)^2 & (216)^2 & (157)^2 & (85)^2 \\ (80)^2 & (105)^2 & (85)^2 & (50)^2 \end{bmatrix}$$

b) Polynomial ~~kernel~~ kernel up to degree d = 2

$$K(x,z) = (x \cdot z + 1)^d$$
$$= (x \cdot z + 1)^2$$

$$K = \begin{bmatrix} (161)^2 & (109)^2 & (117)^2 & (81)^2 \\ (109)^2 & (334)^2 & (217)^2 & (106)^2 \\ (117)^2 & (217)^2 & (158)^2 & (86)^2 \\ (81)^2 & (106)^2 & (86)^2 & (51)^2 \end{bmatrix}$$

c) RBF kernel with $\gamma = 1$

$$K(x,z) = \exp\left(-1 \|x - z\|^2\right)$$
$$= \exp\left(-1 (\|x\|)^2 - 1 (\|z\|)^2 + 2(x \cdot z')\right)$$
$$= \exp\left(-1(\|x\|)^2\right) \cdot \exp\left(-1(\|z\|)^2\right)$$
$$\cdot \exp\left(2 \langle x, z' \rangle\right)$$

continued in last page.

$\|x_1\|^2 = 160$

$\|x_2\|^2 = 333$

$\|x_3\|^2 = 157$

$\|x_4\|^2 = 50.$

$$K = \begin{bmatrix} \exp(-160)\exp(-160)\exp(2(160)) & \cdots \\ \exp(-333)\exp(-160)\exp(2(108)) & \cdots \\ \exp(-157)\exp(-160)\exp(2(116)) & \cdots \\ \exp(-50)\exp(-160)\exp(2(81)) & \cdots \end{bmatrix}$$

5.

The VC-dimension of hypothesis space = 2

Sample complexity grows polynomially in $1/\epsilon$, $1/\delta$.

$$m \geq \frac{1}{\epsilon}\left(4\log_2\left(\frac{2}{\delta}\right) + 8\ VC\text{-}dim(H)\log_2\left(\frac{13}{\epsilon}\right)\right)$$

We can specify a polynomial time algorithm for finding consistent hypothesis:

1. Sort training instances by distance from origin.
2. Set r to be less than the distance to first pos in sorted list.
3. Set $r+a$ to be greater than distance from the last position.