BITS F464: Machine Learning

10 Minimum Description Length (MDL)

Dr. Kamlesh Tiwari
Assistant Professor, Department of CSIS,
BITS Pilani, Pilani Campus, Rajasthan-333031 INDIA

Feb 08, 2021 ONLINE (Campus @ BITS-Pilani Jan-May 2021)

http://ktiwari.in/ml

Representing Database

Can an algorithm help?

- Algorithm can sequentially generate complete observation space
- and can allocate unique identity
- Database as subset of observation could use these identities
- What effect this representation approach has?

Minimum Description Length (MDL)

Kolmogorov complexity of data 1963

Kolmogorov complexity of a data is the length of the shortest computer program that produces it as output.

- Compare 1111111111111111 and 0000000000000000
- Compare 1111111111111111 and 1111111111111111
- Compare 1111111111111111 and 101100110

by Jorma Rissanen in 1978

Goal of statistical inference should be to find *regularity* in the data that can be identified by *ability to compress*

- Learning is viewed as data compression
- One should select a hypothesis that compresses the data most

Minimum Description Length (MDL)

Day	Outlook x_1	Temperature x_2	Humidity x_3	Wind x_4	Play y
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rainy	Mild	High	Weak	Yes
D5	Rainy	Cool	Normal	Weak	Yes
D6	Rainy	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rainy	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rainy	Mild	High	Strong	No

- Each tuple can be represented as set of attribute pairs as E = { {x_1=sunny, x_2=hot, x_3=high, x_4=weak}, {x_1=sunny, x_2=hot, x_3=high, x_4=strong}, ... }
- Total number of different attribute values? are 10
- Each tuple have 4 attribute that could be chosen in at most ${}^{10}C_4$ ways =210
- Class attributes are two so total number of ways are 210×2

Minimum Description Length (MDL)

- Code length in bits $\log(210 \times 2) = 8.715$
- Code length of whole database $L(E) = 8.715 \times 14 = 122.01$

Consider two Hypothesis

H1: [play=yes]
If {*outlook*=overcast}
If {*humidity*=normal, *wind*=weak}

H2: [play=yes]
If {*outlook*=overcast}
If {*humidity*=normal, *wind*=weak}
If {*temperature*=mild, *humidity*=normal}

- Length of Hypothesis $L(H1) = \log({}^{10}C_1) + \log({}^{10}C_2) = 8.81$
- Length of Hypothesis $L(H2) = \log({}^{10}C_1) + \log({}^{10}C_2) + \log({}^{10}C_2) = 14.30$

Minimum Description Length (MDL)

Next we need to encode exceptions: (what is tp, tn, fp, fn)

Outlook x_1	Temperature x_2	Humidity x_3	Wind x_4	Play y	H1	H2
sunny	hot	high	weak	no		
sunny	hot	high	strong	no		
overcast	hot	high	weak	yes	y	y
rain	mild	high	weak	yes		
rain	cool	normal	weak	yes	y	y
rain	cool	normal	strong	no		
overcast	cool	normal	strong	yes	y	y
sunny	mild	high	weak	no		
sunny	cool	normal	weak	yes	y	y
rain	mild	normal	weak	yes	y	y
sunny	mild	normal	strong	yes		y
overcast	mild	high	strong	yes	y	y
overcast	hot	normal	weak	yes	y	y
rain	mild	high	strong	no		

H1

Actual/Predicted	Yes	No
Yes	7	2
No	0	5

H2

Actual/Predicted	Yes	No
Yes	8	1
No	0	5

Minimum Description Length (MDL)

H1

Actual/Predicted	Yes	No
Yes	7	2
No	0	5

H2

Actual/Predicted	Yes	No
Yes	8	1
No	0	5

- $L(E/H1) = \log({}^{7}C_0) + \log({}^{7}C_2) = 4.39$
- $L(E/H2) = \log({}^{8}C_0) + \log({}^{6}C_1) = 2.59$

- Compression of H1: $L(E) - L(H1) - L(E/H1) = 108.81$
- Compression of H2: $L(E) - L(H2) - L(E/H2) = 105.12$

Result shows that H1 has better compression than H2.

So, we should stop generating more rules and deliver H1 as good model

Thank You!

Thank you very much for your attention!

Queries ? Ref