Modul : Supervised Learning

Naive Bayes

Nur ULFA Maulidevi

KK IF - Teknik Informatika- STEI ITB

Inteligensi Buatan
(Artificial Intelligence)

Naive Bayes

model probabilitas/peluang

Supervised Learning

Probabilistic Classifier

Target function: class from finite set V

Hypothesis: Probability Model of $P(v_j)$ and $P(a_i|v_j)$

Unseen data prediction: Find the maximum $P(v_j \mid a_1, a_2, ..., a_n)$

EDUNEX ITB

Learning Algorithm

Measures the frequency of every attribute value for certain class, and the frequency of every class

Determine the probability of $P(a_i|v_j)$

Determine the probability of $P(v_j)$

Hypothesis: Probability Model

EDUNEX ITB

Classification (Predict Unseen Data)

Compute the projection of attribute probability of query (unseen data) for every class $\square \prod_i P(a_i \mid v_j)$

Multiply the result with probability of each class to get $\square$
$$P(v_j \mid a_1, a_2, ..., a_n) = P(v_j) \cdot \prod_i P(a_i \mid v_j)$$

Prediction: Class with the maximum $P(v_j \mid a_1, a_2, ..., a_n)$

EDUNEX ITB

Example: Play Tennis Dataset

outlook	temp.	humidity	windy	play
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	hot	high	false	yes
rainy	mild	high	false	yes
rainy	cool	normal	false	yes
rainy	cool	normal	true	no
overcast	cool	normal	true	yes

outlook	temp.	humidity	windy	play
sunny	mild	high	false	no
sunny	cool	normal	false	yes
rainy	mild	normal	false	yes
sunny	mild	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes
rainy	mild	high	true	no

outlook	yes	no	temperature	yes	no	humidity	yes	no	windy	yes	no	play	yes	no
sunny	2	3	hot	2	2	high	3	4	false	6	2		9	5
overcast	4	0	mild	4	2	normal	6	1	true	3	3			
rainy	3	2	cool	3	1									

Frequency of (sunny|yes)

Frequency of (sunny|no)

Frequency of class 'yes' 9

Frequency of class 'no' 5

Example: Play Tennis Dataset

Frequency

outlook	yes	no	temperature	yes	no	humidity	yes	no	windy	yes	no	play	yes	no
sunny	2	3	hot	2	2	high	3	4	false	6	2		9	5
overcast	4	0	mild	4	2	normal	6	1	true	3	3			
rainy	3	2	cool	3	1									

$P(a_i|v_j)$

$P(v_j)$

Probability Model

outlook	yes	no	temperature	yes	no	humidity	yes	no	windy	yes	no
sunny	2/9	3/5	hot	2/9	2/5	high	3/9	4/5	false	6/9	2/5
overcast	4/9	0/5	mild	4/9	2/5	normal	6/9	1/5	true	3/9	3/5
rainy	3/9	2/5	cool	3/9	1/5						

play	
yes	no
9/14	5/14

Classify New Instance: <Sunny, Cool, High, True>

$P(a_i|v_j)$

	outlook		temperature		humidity		windy	
	yes	no	yes	no	yes	no	yes	no
sunny	2/9	3/5	hot 2/9	2/5	high 3/9	4/5	false 6/9	2/5
overcast	4/9	0/5	mild 4/9	2/5	normal 6/9	1/5	true 3/9	3/5
rainy	3/9	2/5	cool 3/9	1/5				

$P(v_j)$

play	
yes	no
9/14	5/14

$$P(v_j \mid a_1, a_2, ..., a_n) = P(v_j) \cdot \prod_i P(a_i \mid v_j)$$

P(yes|sunny, cool, high, true)

= P(yes). P(sunny|yes).P(cool|yes).P(high|yes).P(true|yes)

= 9/14 . 2/9 . 3/9 . 3/9 . 3/9 = 0.0053

P(no |sunny, cool, high, true)

= P(no). P(sunny|no).P(cool| no).P(high| no).P(true| no)

= 5/14 . 3/5 . 1/5 . 4/5 . 3/5 = 0.0206

EDUNEX ITB

THANK YOU