Lec 9 Classification 2

Problems of KNN
1) require large data storage
2) demanding task for calculating distance
3) slow prediction result

Logistic Function
a simple arithmetic equation

Person	Height	Weight	Gender
P1	0.625	0.875	M
P2	0	0	F
P3	0.25	0.375	M
P4	1	1	M
P5	0.4583	0.6667	??

there are two variables to consider
H : height
W : weight

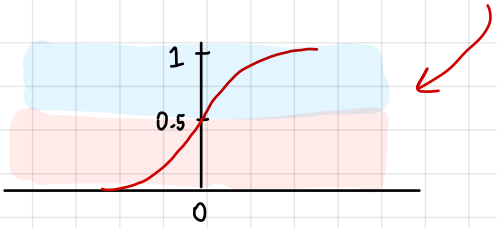
The result = male / female
male : 1
female : 0

The simple equation : $H + W \geq 0.5 \rightarrow$ male

more complex equation : $w_h H + w_w W + w_0 \geq 0.5$
w_h, w_w serve as weight to indicate significance
w_0 serve as bias to adjust the threshold
0.5 is the current threshold

The logistic function is in form of $\frac{1}{1+e^{-t}}$ which the value $\in (0, 1]$

adopt it → $\frac{1}{1+e^{-(w_h H + w_w W + w_0)}} \geq 0.5$



To train logisfic function, w_h, w_w and w_0 need to be carefully selected
- recursive method
till the model fit the data trained [can correctly classified all]

Loss function
to measure the fitness of the model
Ground truth : male = 1 , female = 0

Person	Height	Weight	Gender
P1	0.625	0.875	M

Loss function $(Y_{output} - Y)^2 = 0.0723$

try to change Youtput till loss function is minimized.

$w_h = 1$
$w_w = 1$
$w_0 = -0.5$

$\frac{1}{1+e^{-(0.625+0.875-0.5)}} = 0.731$

Y_{output} : 0.731
Y_{ground} : 1

How to minimize loss function ?

Gradient descent algorithm

find a value that makes function smallest

1) random w_0, w_w and w_0
2) find Y_{output}
   - update $w_i = w_i + \Delta w_i$
     while $\Delta w_i = 2 * \alpha (Y - Y_{output}) \frac{\partial Y_{output}}{\partial w_i}$
3) repeat till no update

recursive function