

Họ tên: Mai Thuý Ngọc

Lớp: 58K2-CNTT

BÀI TẬP THỰC HÀNH 3: K-Nearest Neighbors

Câu 1: Hãy sử dụng giải thuật KNN để xác định thể loại cho các phim. Với $K=1, 3, 5$.

Tên phim	Số lượng cú đá	Số lượng nụ hôn	Loại phim
Californiaman	3	104	Tình cảm
He isn't really into dudes	2	100	Tình cảm
Beautiful Woman	1	81	Tình cảm
Kevin Longblade	101	10	Hành động
Robo Slayer 3000	99	5	Hành động
Amped II	98	2	Hành động
Kiss kick	18	90	?
Last blood	23	6	?
Lucky men	8	10	?

Gọi d là euclidean distance

Kí hiệu:

A_1 = Californiaman

A_2 = He isn't really into dudes

A_3 = Beautiful Woman

A_4 = Kevin Longblade

A_5 = Robo Slayer 3000

A_6 = Amped II

❖ Kiss kick

$$d_{(\text{Kiss kick}, A_1)} = \sqrt{(18 - 3)^2 + (90 - 104)^2} = 20.518$$

$$d_{(\text{Kiss kick}, A_2)} = \sqrt{(18 - 2)^2 + (90 - 100)^2} = 18.867$$

$$d_{(\text{Kiss kick}, A_3)} = \sqrt{(18 - 1)^2 + (90 - 81)^2} = 19.235$$

$$d_{(\text{Kiss kick}, A_4)} = \sqrt{(18 - 101)^2 + (90 - 10)^2} = 115.277$$

$$d_{(\text{Kiss kick}, A_5)} = \sqrt{(18 - 99)^2 + (90 - 5)^2} = 117.413$$

$$d_{(\text{Kiss kick}, A_6)} = \sqrt{(18 - 98)^2 + (90 - 2)^2} = 118.928$$

+ Giả sử $K = 1$ thì sẽ có một láng giềng gần nhất là He isn't really into dudes (A2). Theo thuật toán KNN sẽ lấy loại phim nào trong các láng giềng gần nhất để làm loại phim cho phim cần được xác định loại ($d_{(\text{Kiss kick}, A2)}$ nhỏ nhất) \Rightarrow Phim cần xác định thuộc thể loại phim Tình cảm.

+ Giả sử $K = 3$ thì sẽ có 3 láng giềng gần nhất là He isn't really into dudes (A2), Beautiful Woman (A3) và Californiaman (A1). Vì trong 3 láng giềng gần nhất đều có 3 phim đều thuộc thể loại phim Tình cảm \Rightarrow Phim cần xác định thuộc thể loại phim Tình cảm.

+ Giả sử $K = 5$ thì sẽ có 5 láng giềng gần nhất là He isn't really into dudes (A2), Beautiful Woman (A3), Californiaman (A1), Kevin Longblade (A4) và Robo Slayer 3000 (A5). Vì trong 5 láng giềng gần nhất có 3 láng giềng thuộc thể loại phim Tình cảm và có 2 láng giềng thuộc thể loại phim Hành động \Rightarrow Phim cần xác định thuộc thể loại phim Tình cảm.

❖ Last blood

$$d_{(\text{Last blood}, A1)} = \sqrt{(23 - 3)^2 + (6 - 104)^2} = 100.019$$

$$d_{(\text{Last blood}, A2)} = \sqrt{(23 - 2)^2 + (6 - 100)^2} = 96.317$$

$$d_{(\text{Last blood}, A3)} = \sqrt{(23 - 1)^2 + (6 - 81)^2} = 78.160$$

$$d_{(\text{Last blood}, A4)} = \sqrt{(23 - 101)^2 + (6 - 10)^2} = 78.102$$

$$d_{(\text{Last blood}, A5)} = \sqrt{(23 - 99)^2 + (6 - 5)^2} = 76.006$$

$$d_{(\text{Last blood}, A6)} = \sqrt{(23 - 98)^2 + (6 - 2)^2} = 75.106$$

+ Giả sử $K = 1$ thì sẽ có một láng giềng gần nhất là Amped II (A6). Theo thuật toán KNN sẽ lấy loại phim nào trong các láng giềng gần nhất để làm loại phim cho phim cần được xác định loại ($d_{(\text{Last blood}, A6)}$ nhỏ nhất) \Rightarrow Phim cần xác định thuộc thể loại phim Hành động.

+ Giả sử $K = 3$ thì sẽ có 3 láng giềng gần nhất là Amped II (A6), Robo Slayer 3000 (A5) và Kevin Longblade (A4). Vì trong 3 láng giềng gần nhất đều có 3 phim đều thuộc thể loại phim Hành động \Rightarrow Phim cần xác định thuộc thể loại phim Hành động.

+ Giả sử $K = 5$ thì sẽ có 5 láng giềng gần nhất là Amped II (A6), Robo Slayer 3000 (A5) và Kevin Longblade (A4), Beautiful Woman (A3) và He isn't really into dudes (A2). Vì trong 5 láng giềng gần nhất có 3 láng giềng thuộc thể loại phim Hành động và có 2 láng giềng thuộc thể loại phim Tình cảm \Rightarrow Phim cần xác định thuộc thể loại phim Hành động.

❖ Lucky men

$$d_{(\text{Lucky men}, A1)} = \sqrt{(8 - 3)^2 + (10 - 104)^2} = 94.132$$

$$d_{(\text{Lucky men}, A2)} = \sqrt{(8 - 2)^2 + (10 - 100)^2} = 90.199$$

$$d_{(\text{Lucky men}, A3)} = \sqrt{(8 - 1)^2 + (10 - 81)^2} = 71.344$$

$$d_{(\text{Lucky men}, A4)} = \sqrt{(8 - 101)^2 + (10 - 10)^2} = 93$$

$$d_{(\text{Lucky men}, A5)} = \sqrt{(8 - 99)^2 + (10 - 5)^2} = 91.137$$

$$d_{(\text{Lucky men}, A6)} = \sqrt{(8 - 98)^2 + (10 - 2)^2} = 90.354$$

+ Giả sử $K = 1$ thì sẽ có một láng giềng gần nhất là Beautiful Woman (A3). Theo thuật toán KNN sẽ lấy loại phim nào trong các láng giềng gần nhất để làm loại phim cho phim cần được xác định loại ($d_{(\text{Lucky men}, A3)}$ nhỏ nhất) \Rightarrow Phim cần xác định thuộc thể loại phim Tình cảm.

+ Giả sử $K = 3$ thì sẽ có 3 láng giềng gần nhất là Beautiful Woman (A3), He isn't really into dudes (A2) và Amped II (A6). Vì trong 3 láng giềng gần nhất có 2 láng giềng thuộc thể loại phim Tình cảm và 1 láng giềng thuộc thể loại phim Hành động \Rightarrow Phim cần xác định thuộc thể loại phim Tình cảm.

+ Giả sử $K = 5$ thì sẽ có 5 láng giềng gần nhất là Beautiful Woman (A3), He isn't really into dudes (A2) và Amped II (A6), Robo Slayer 3000 (A5) và Kevin Longblade (A4). Vì trong 5 láng giềng gần nhất có 3 láng giềng thuộc thể loại phim Hành động và có 2 láng giềng thuộc thể loại phim Tình cảm \Rightarrow Phim cần xác định thuộc thể loại phim Hành động.

Kết luận:

TH: $K = 1$ và $K = 3$

Tên phim	Số lượng cú đá	Số lượng nụ hôn	Loại phim
Californiaman	3	104	Tình cảm
He isn't really into dudes	2	100	Tình cảm
Beautiful Woman	1	81	Tình cảm
Kevin Longblade	101	10	Hành động
Robo Slayer 3000	99	5	Hành động
Amped II	98	2	Hành động
Kiss kick	18	90	Tình cảm
Last blood	23	6	Hành động
Lucky men	8	10	Tình cảm

TH: K = 5

Tên phim	Số lượng cú đá	Số lượng nụ hôn	Loại phim
Californiaman	3	104	Tình cảm
He isn't really into dudes	2	100	Tình cảm
Beautiful Woman	1	81	Tình cảm
Kevin Longblade	101	10	Hành động
Robo Slayer 3000	99	5	Hành động
Amped II	98	2	Hành động
Kiss kick	18	90	Tình cảm
Last blood	23	6	Hành động
Lucky men	8	10	Hành động