

Classification + KNN



Week 05 - Day 01

Yes, I know the meaning of KNN in Hokkien!



Classification

Regression = predicting a number

Classification = predicting a class

Pregnant or not?

Apple, orange or pineapple?

Will she churn or not?

Marry him/her or run away?

Binary vs. Multiclass

Example

tinder



Sean Rad, 29 

Verified

 Like Me On Tinder

tinder



Rosette, 32

Vice President, Global Communications & Bra...

 Like Me On Tinder

Should I date him/her?

Previous experiences

+

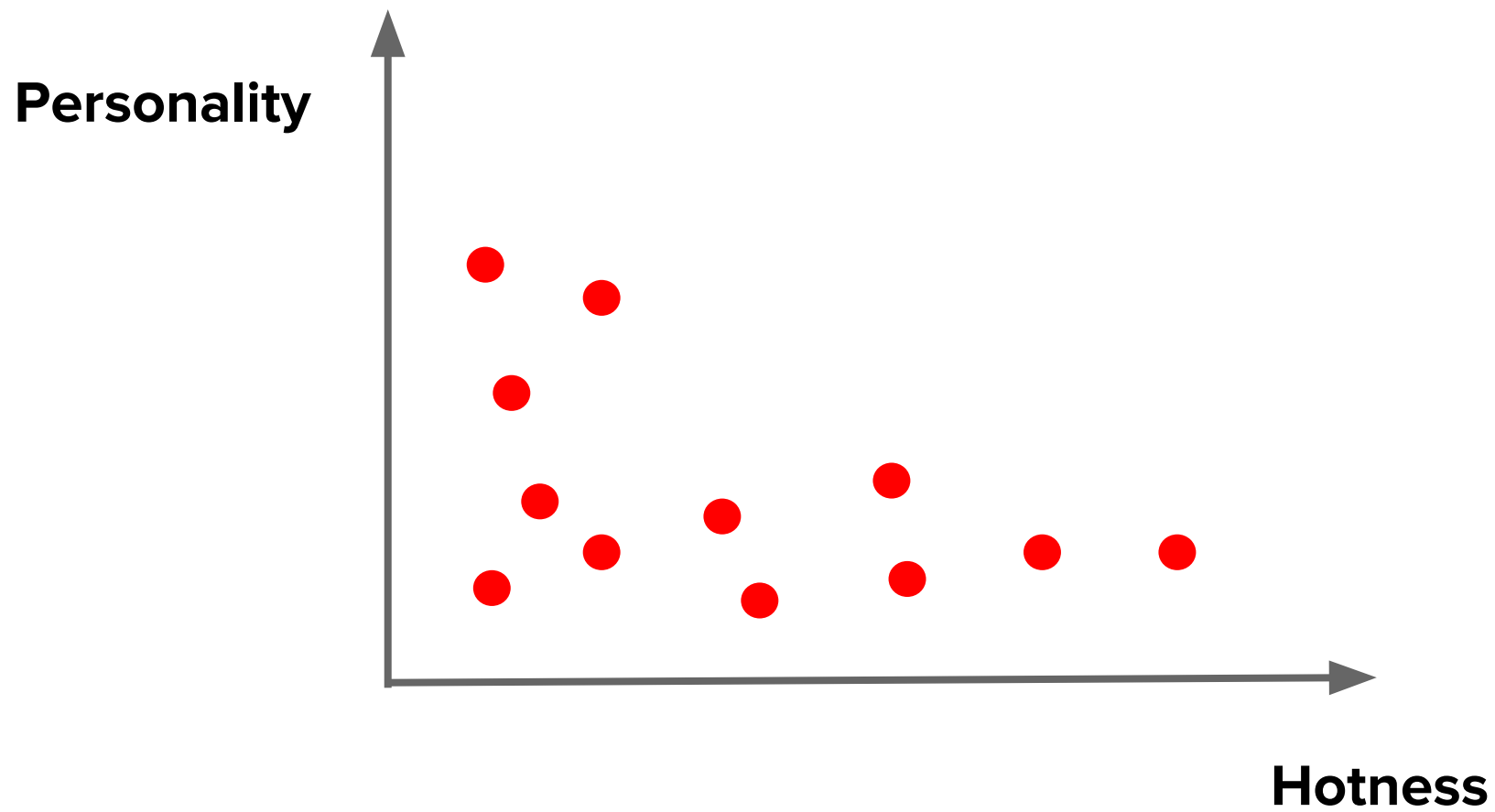
Experiences from friends

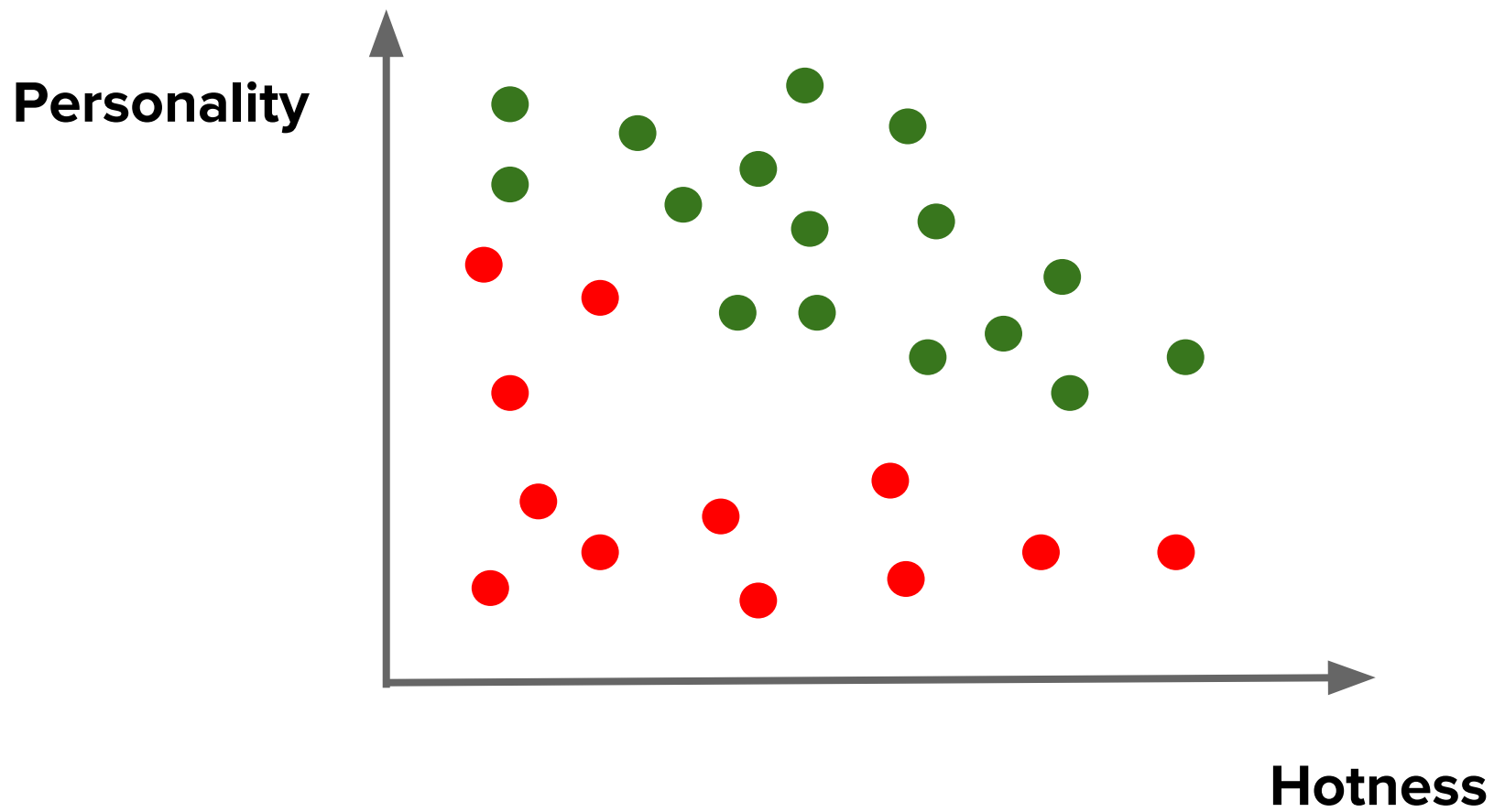
Personality

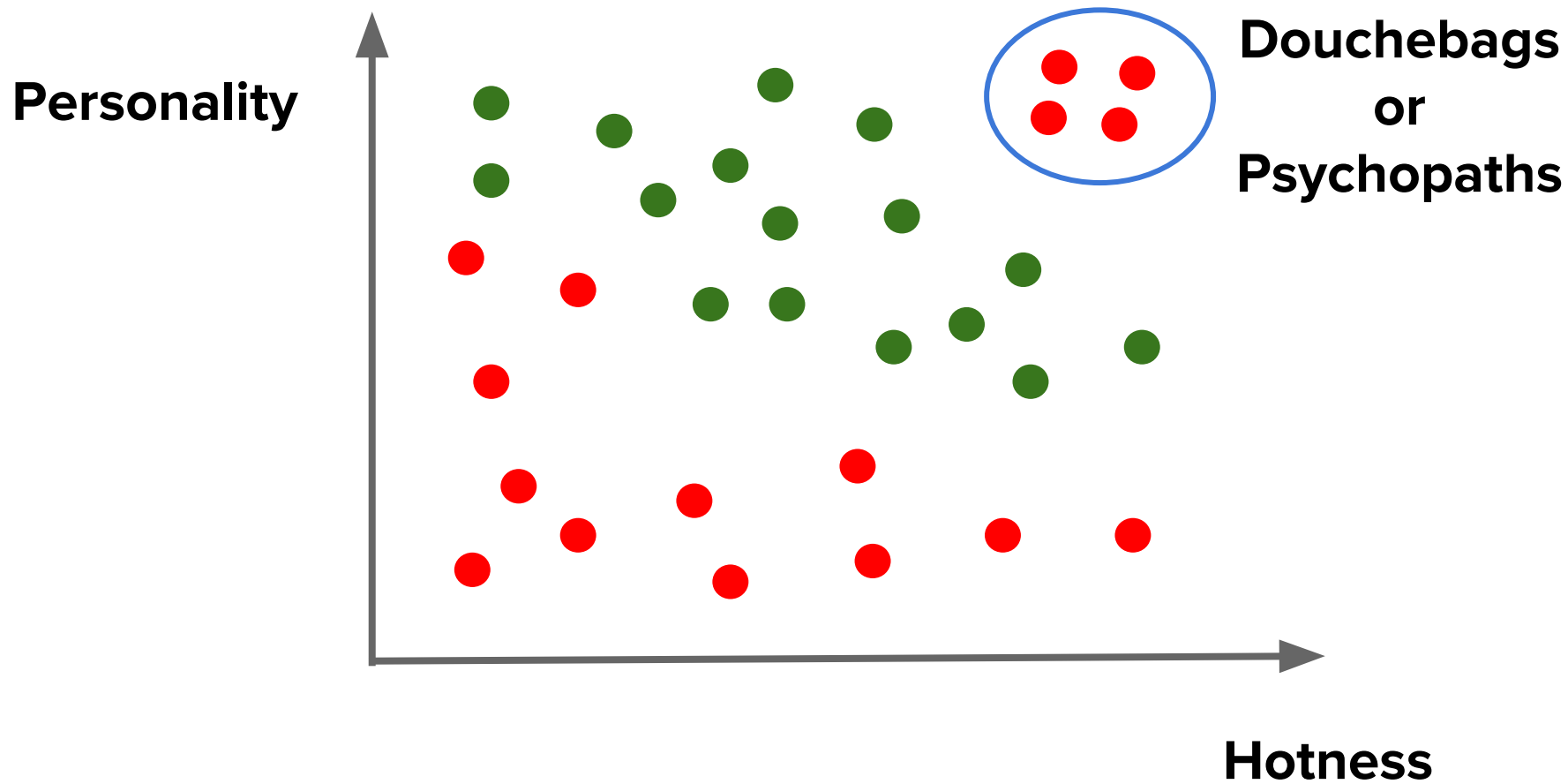


A 2D coordinate system is shown with a vertical axis and a horizontal axis. The vertical axis is labeled 'Personality' at its top, and the horizontal axis is labeled 'Hotness' at its right end. Both axes are represented by dark gray lines with arrowheads at their respective ends. The axes meet at an origin point in the bottom-left corner of the plot area.

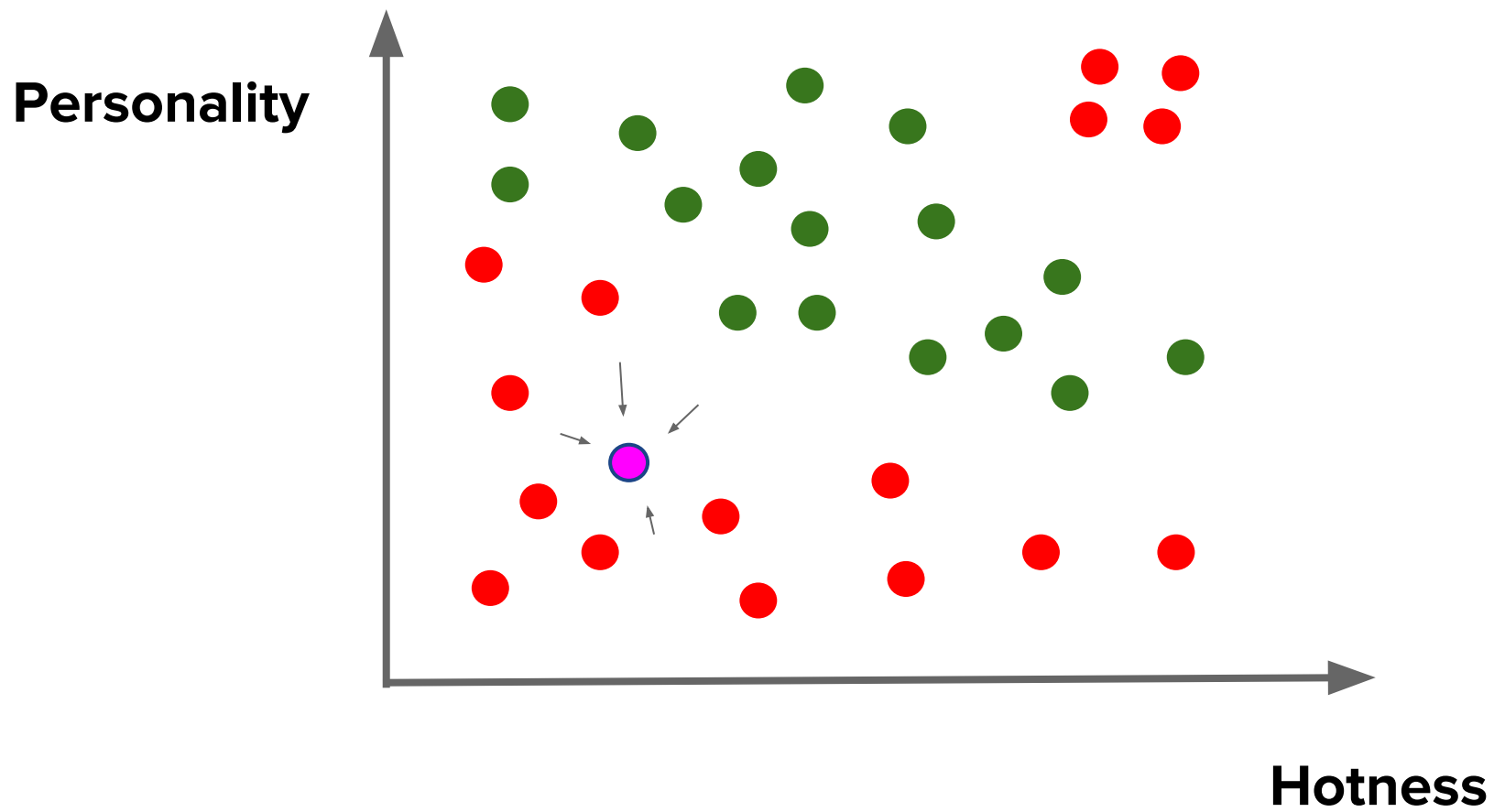
Hotness

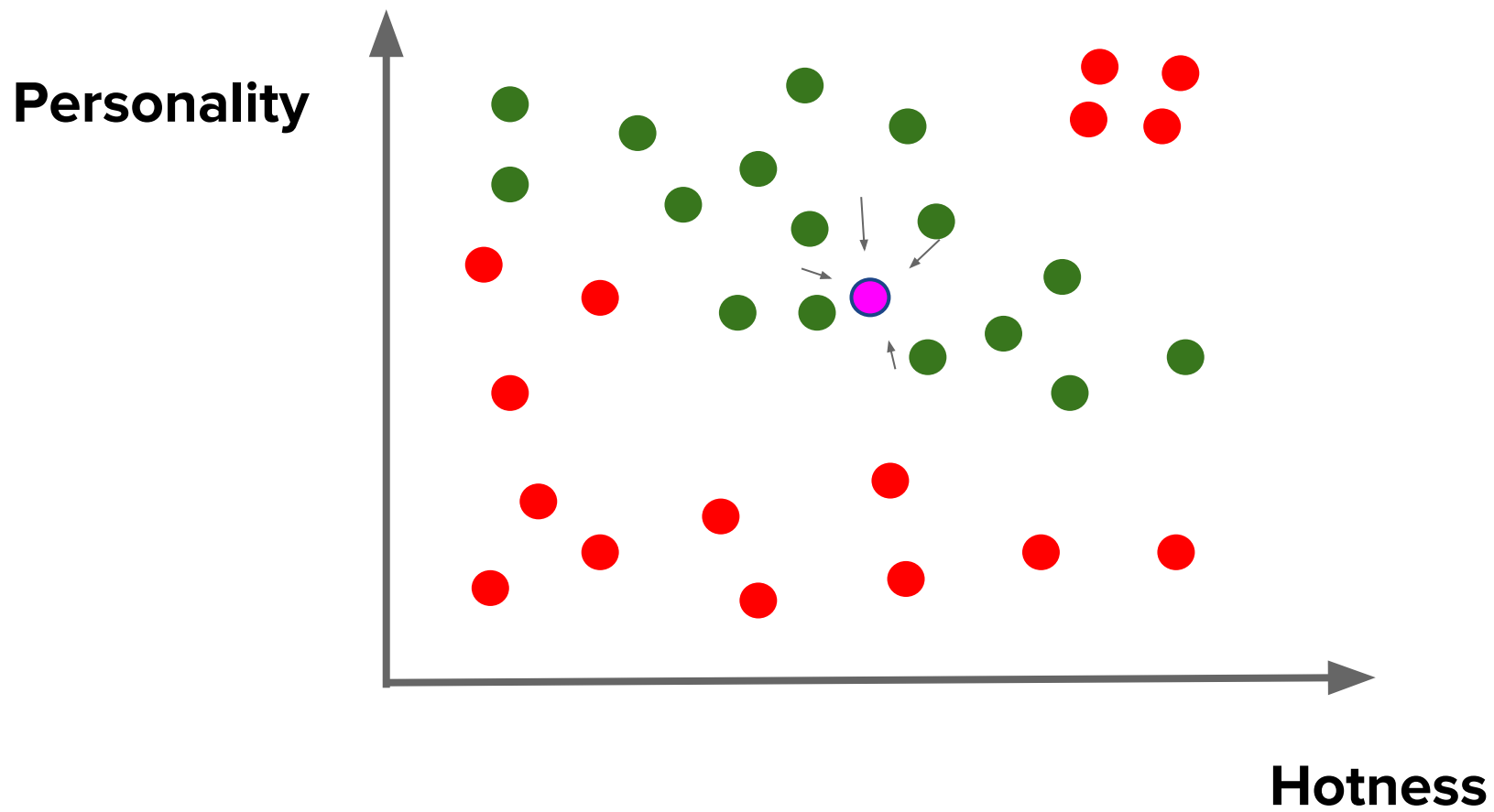


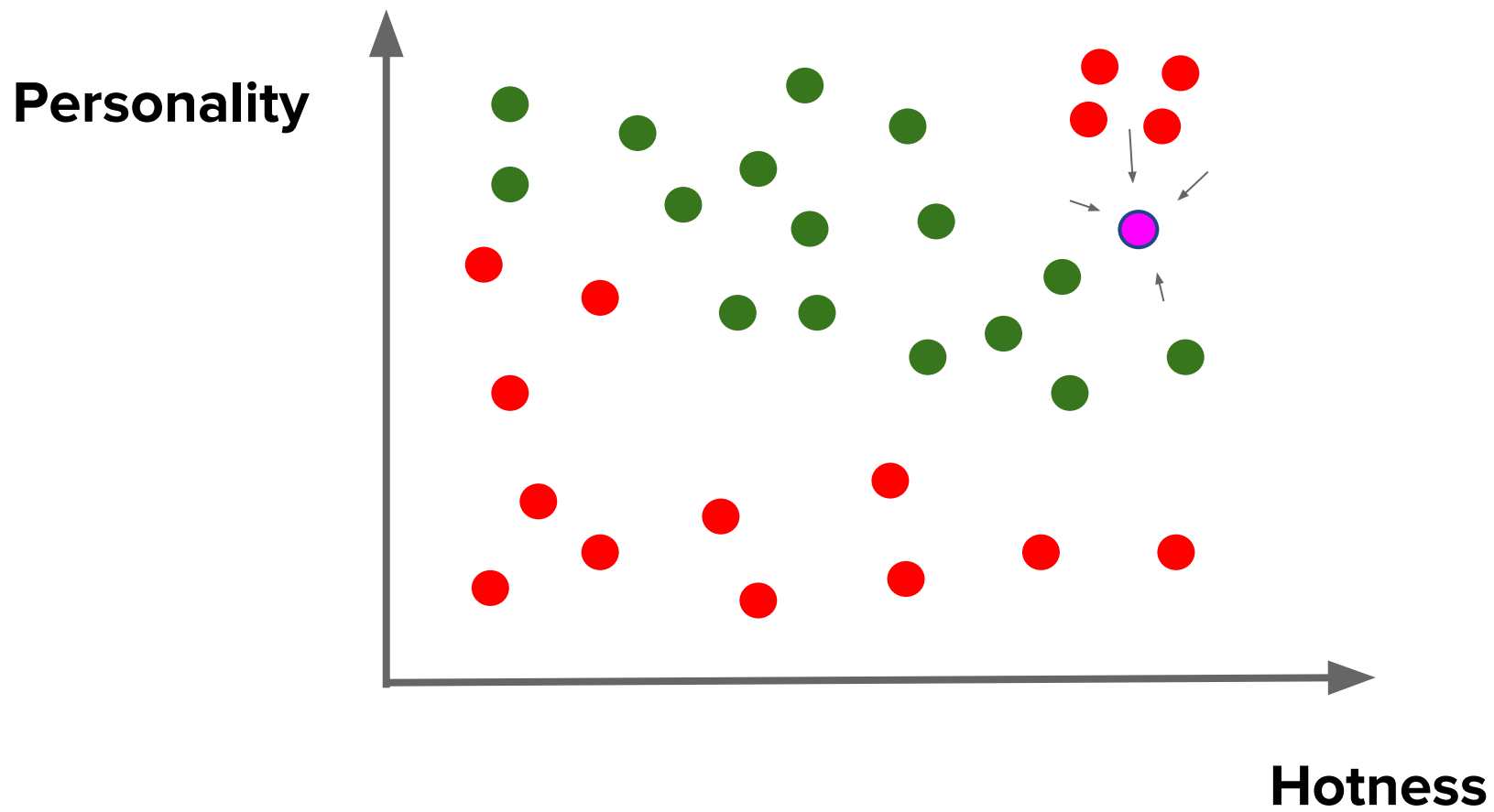




New point - prediction?







Possible solution?

Baseline

What's a simple baseline?

Baseline: most frequent class

- KNN -

K-Nearest Neighbors

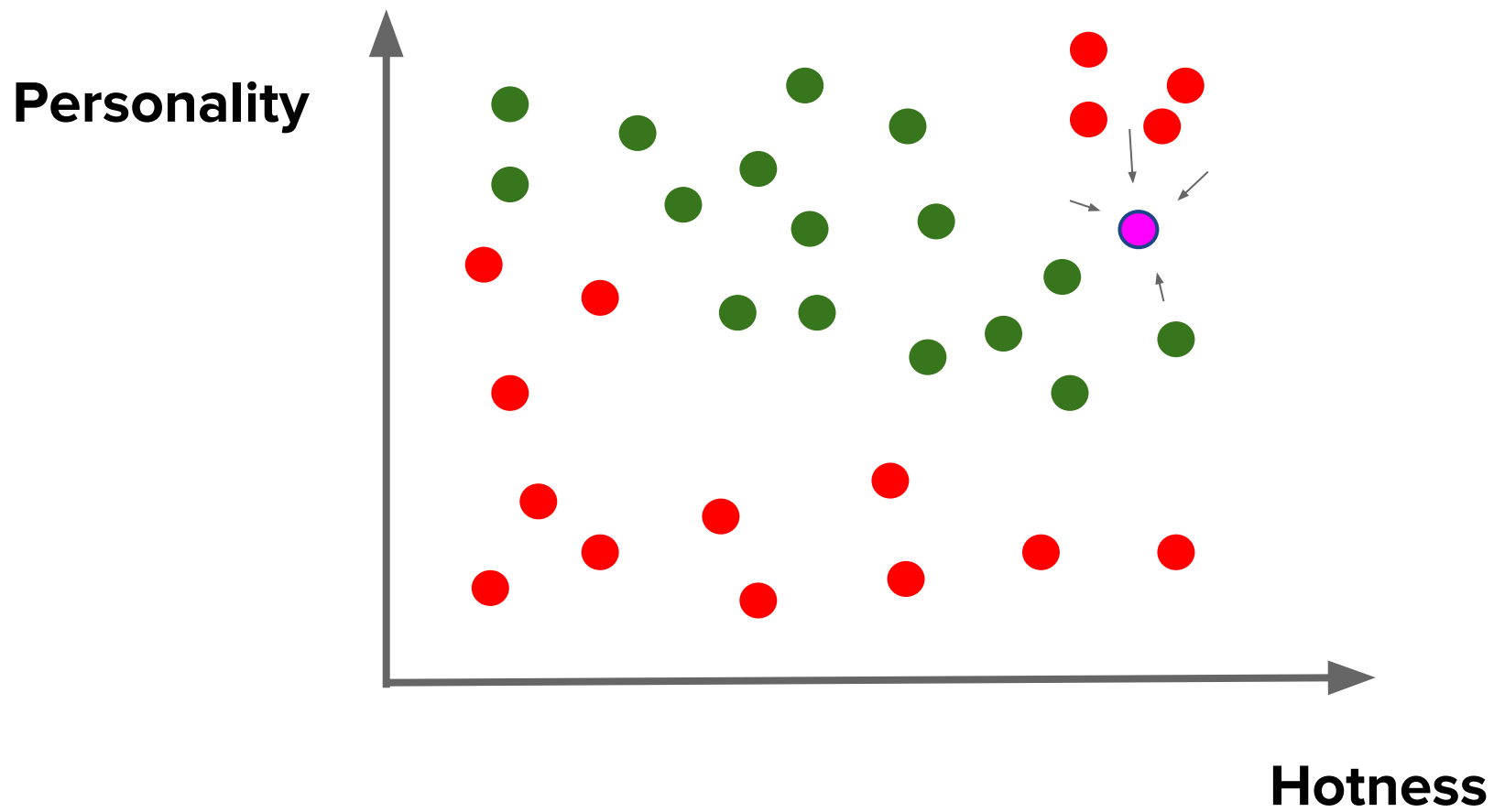
Predicting a new point:

Predicting a new point:

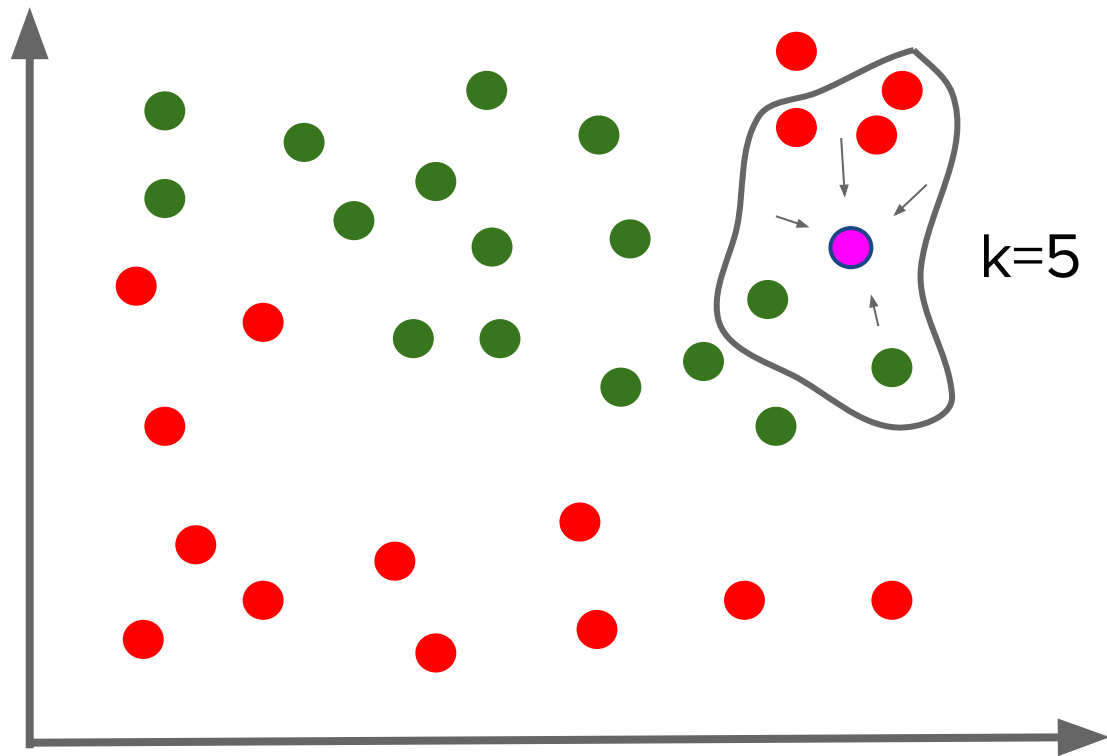
- 1) Calculate the distance to all other points

Predicting a new point:

- 1) Calculate the distance to all other points
- 2) Find the k-nearest neighbors



Personality



$k=5$

Hotness

**What's the best k ?
(2,5,10,100,1000)**

Small K - what happens?

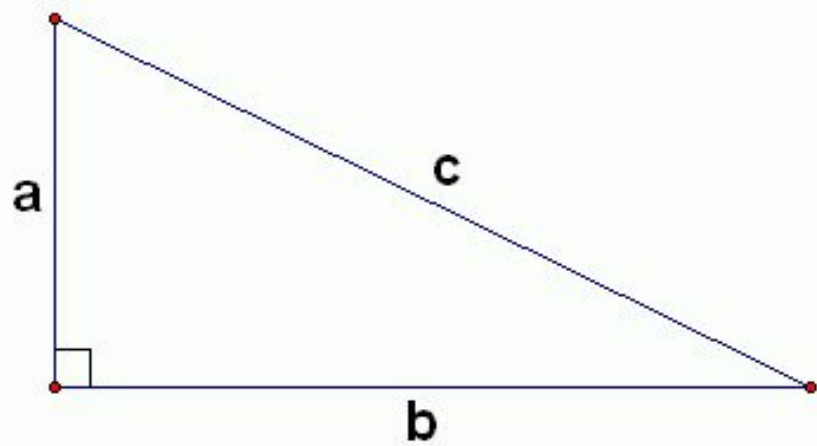
Small K - Noise + variance

Big K - what happens?

Big K - too “general”

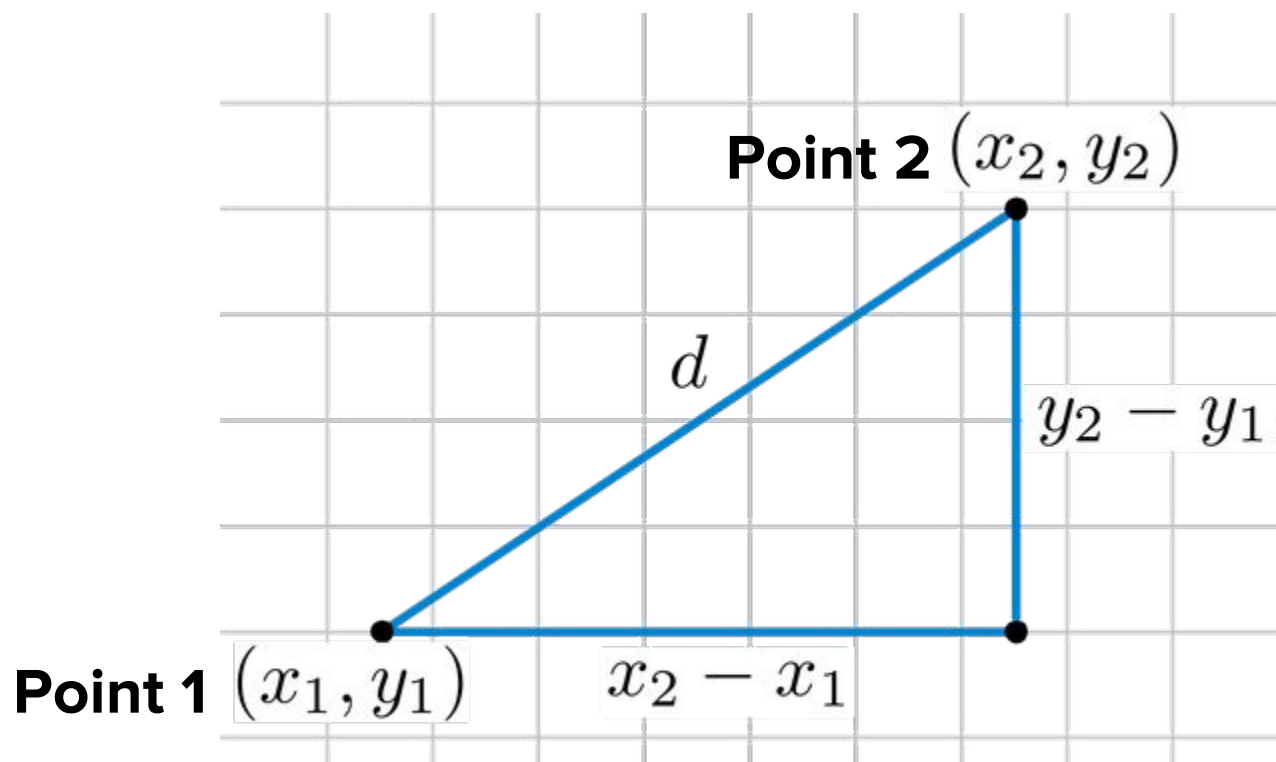
Best K? Tuning!

Euclidean Distance



$$a^2 + b^2 = c^2$$

$$\text{Euclidean distance} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_1)^2}$$



It can be applied to n -dimensional vectors

There are other distances you can use
(a parameter to tune)

**Features Scaling:
Yes or not?**

Features scaling: Yes or Not?

Features scaling: Yes!

**Is KNN a linear
model?**

No...and this is good!

The model

Linear regression = one formula

KNN = ???

Linear regression = one formula

KNN = the entire dataset

Training Vs. Prediction

Linear regression:

Training time: fast/slow?

Prediction time: fast/slow?

Linear regression:

Training time: **slow**

Prediction time: **fast**

KNN:

Training time: fast/slow?

Prediction time: fast/slow?

KNN:

Training time: **fast**

Prediction time: **slow**

Assumptions





Summary

1. Classification: predicting a class

1. Classification: predicting a class
2. KNN: simple algorithm based on distances

1. Classification: predicting a class
2. KNN: simple algorithm based on distances
3. K needs to be tuned

1. Classification: predicting a class
2. KNN: simple algorithm based on distances
3. K needs to be tuned
4. Fast for training, slow for predictions

1. Classification: predicting a class
2. KNN: simple algorithm based on distances
3. K needs to be tuned
4. Fast for training, slow for predictions
5. Scale your features!

1. Classification: predicting a class
2. KNN: simple algorithm based on distances
3. K needs to be tuned
4. Fast for training, slow for predictions
5. Scale your features!
6. KNN doesn't usually perform well