

Classification + KNN



Week 05 - Day 01

Classification

Regression = predicting a number

Classification = predicting ???

Regression = predicting a number

Classification = predicting a class

Pregnant or not?

Apple or orange?

Will she churn or not?

Marry him/her vs. 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

+

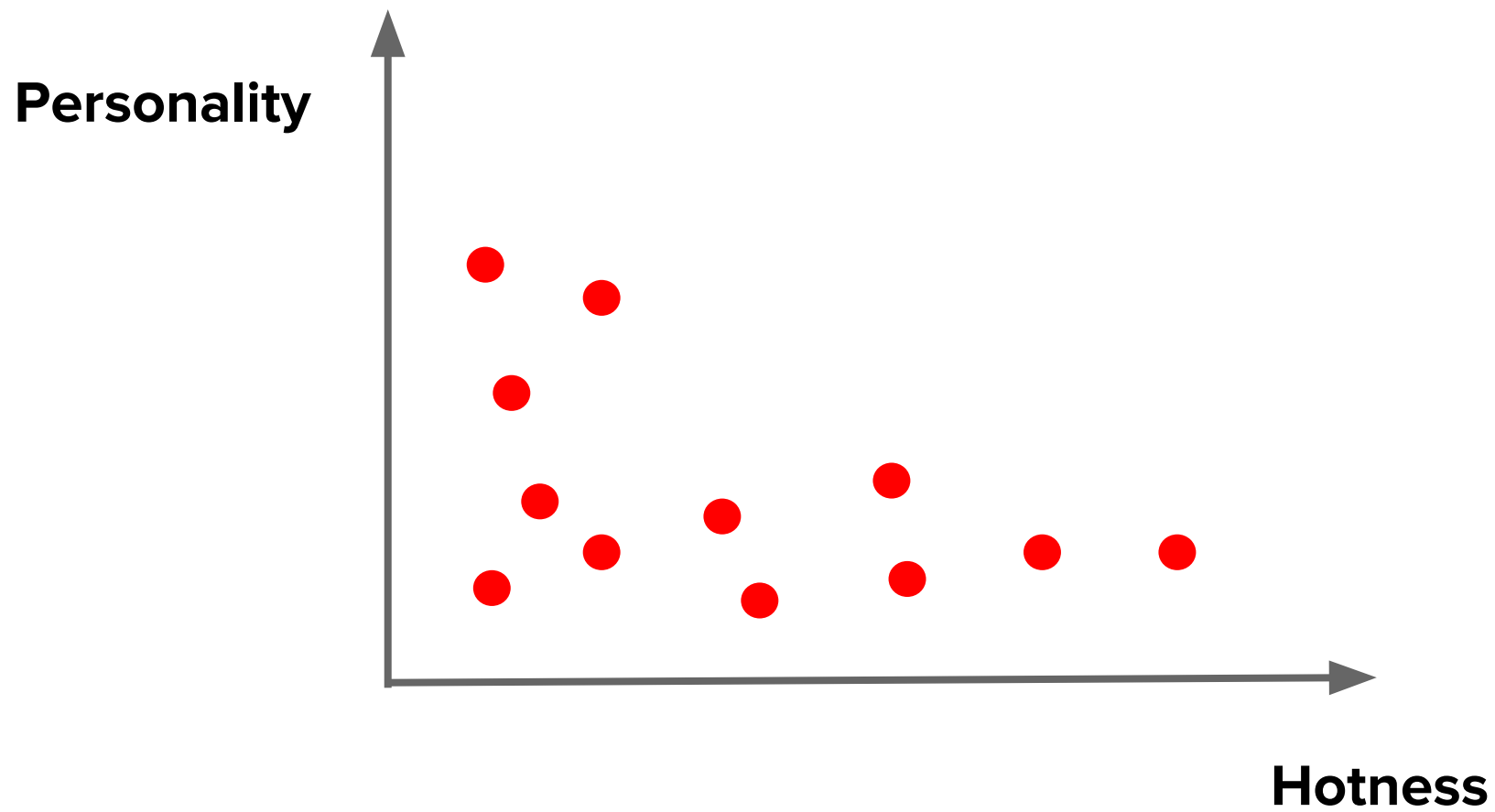
Experiences from friends

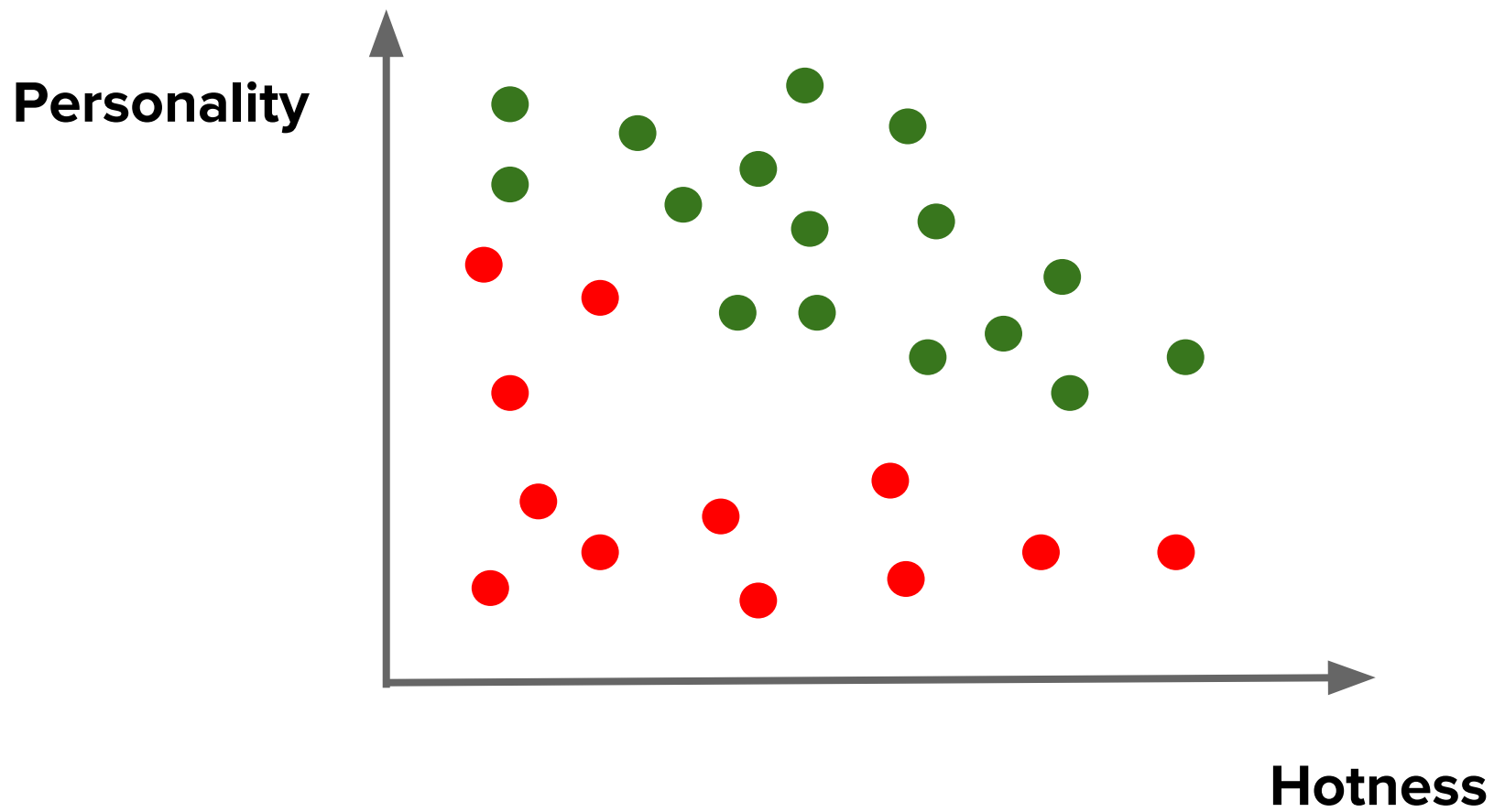
Personality

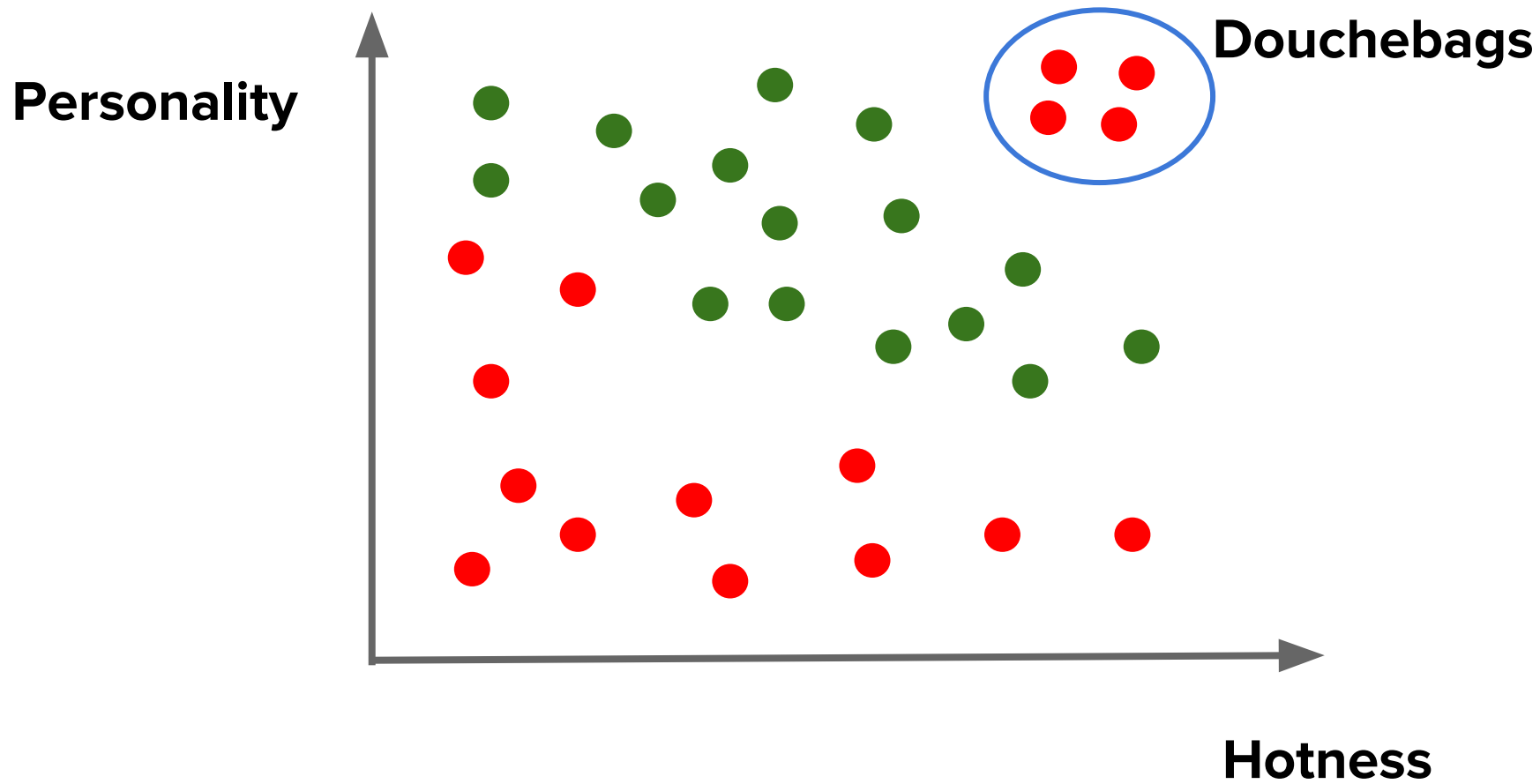


A 2D coordinate system is shown with a vertical y-axis and a horizontal x-axis. Both axes are represented by dark gray lines with arrowheads at their ends. The y-axis is labeled 'Personality' at its top, and the x-axis is labeled 'Hotness' at its right end. The axes meet at an origin in the bottom-left corner, forming an L-shape. The background is a light gray grid.

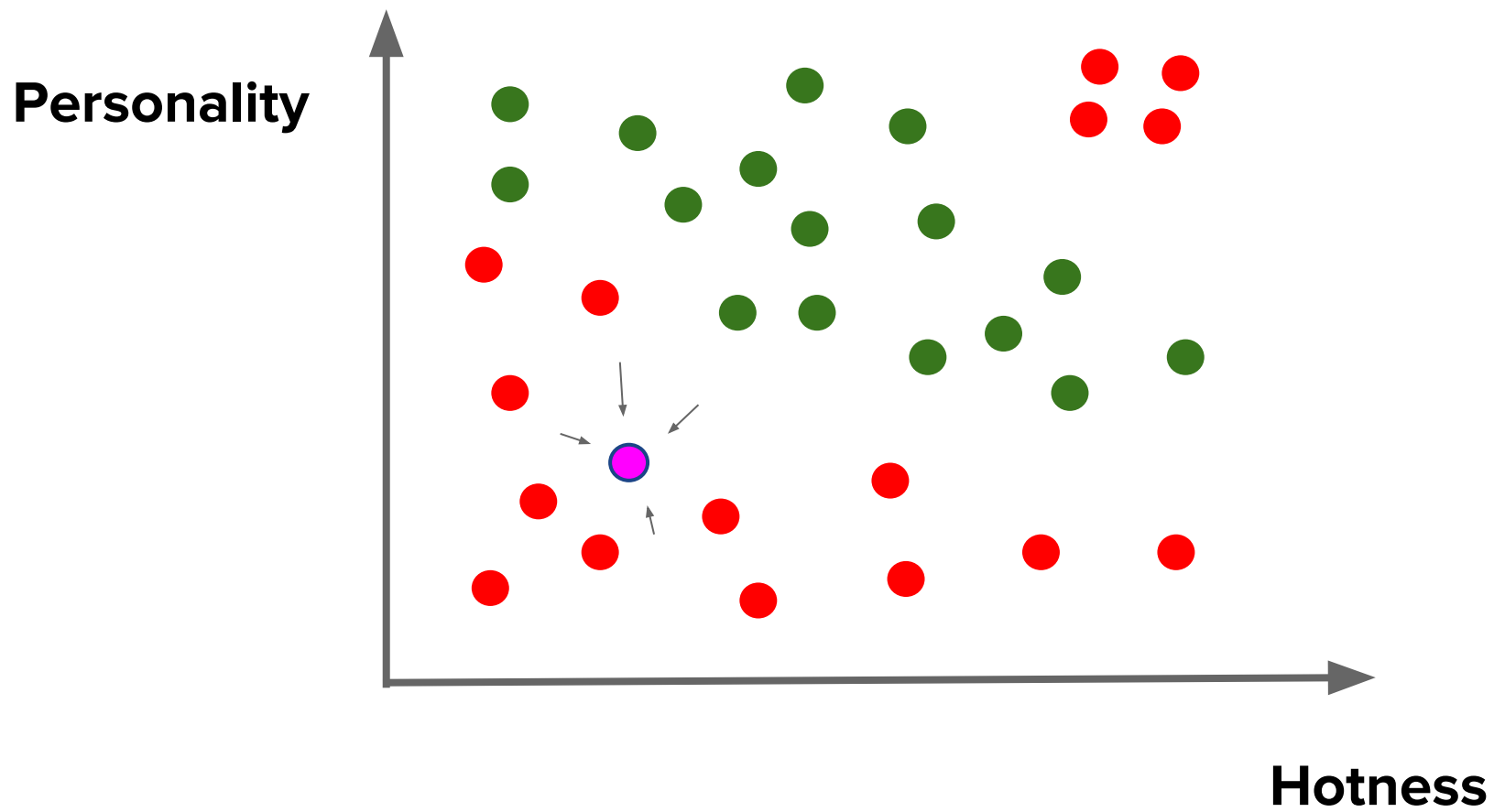
Hotness

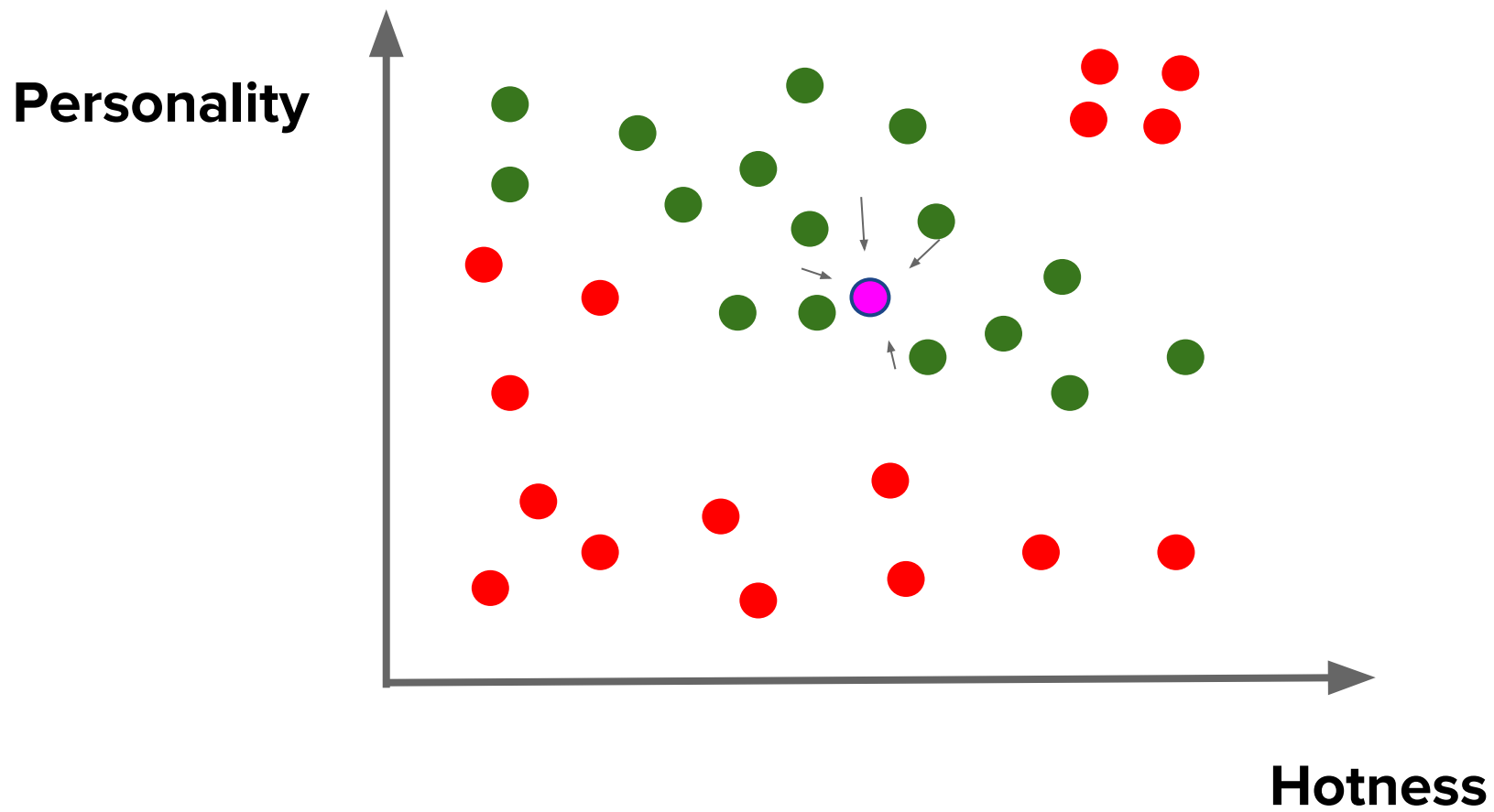


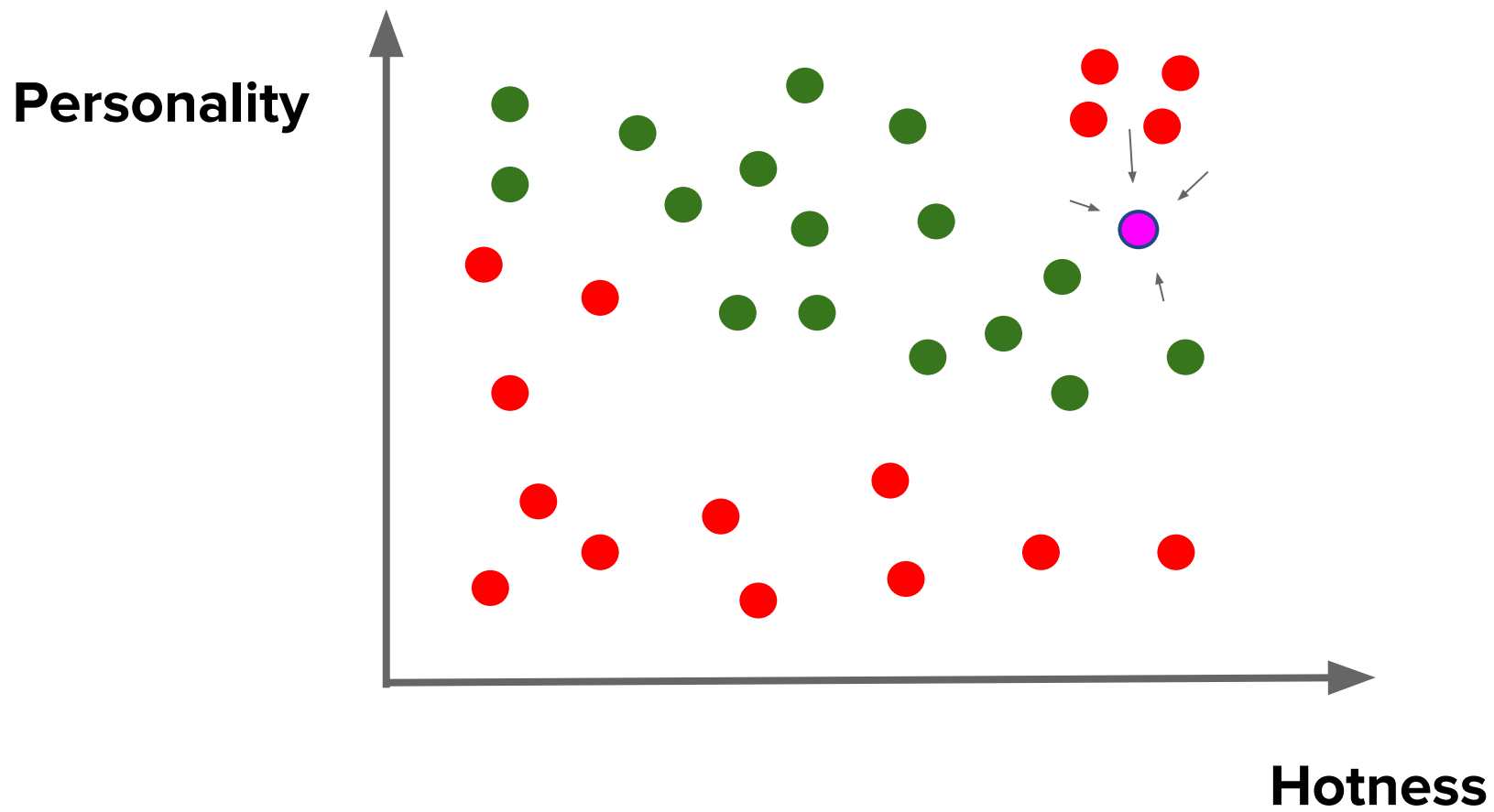




New point - prediction?







Possible solution?

Baseline

What's a simple baseline?

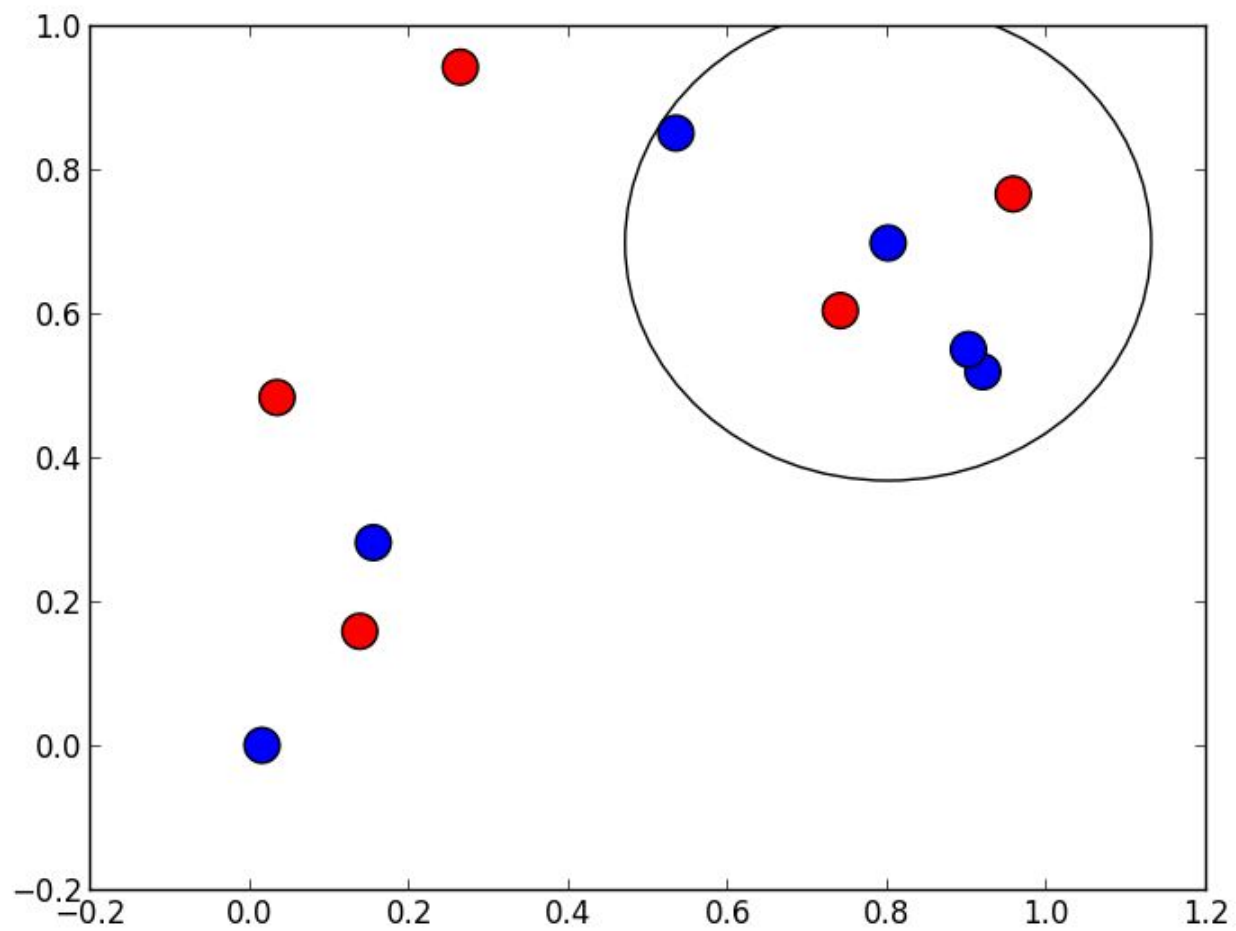
Baseline: most frequent class

- KNN -

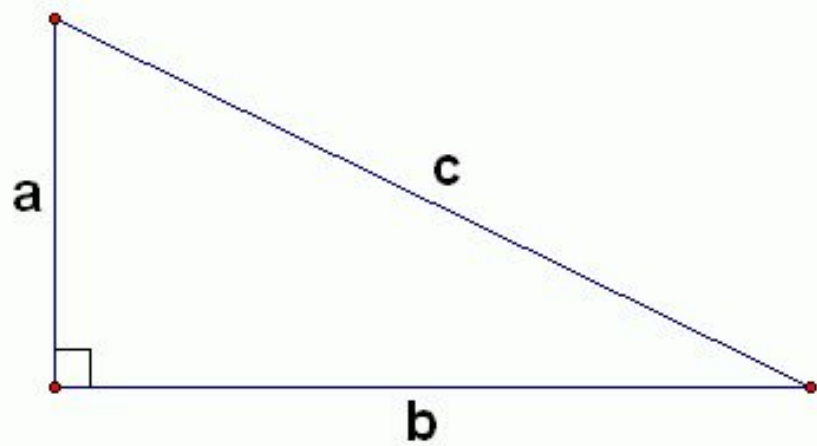
K-Nearest Neighbors

Predicting a new point:

- 1) Calculate the distance to all other points
- 2) Find the k-nearest neighbors
- 3) Return the most frequent class

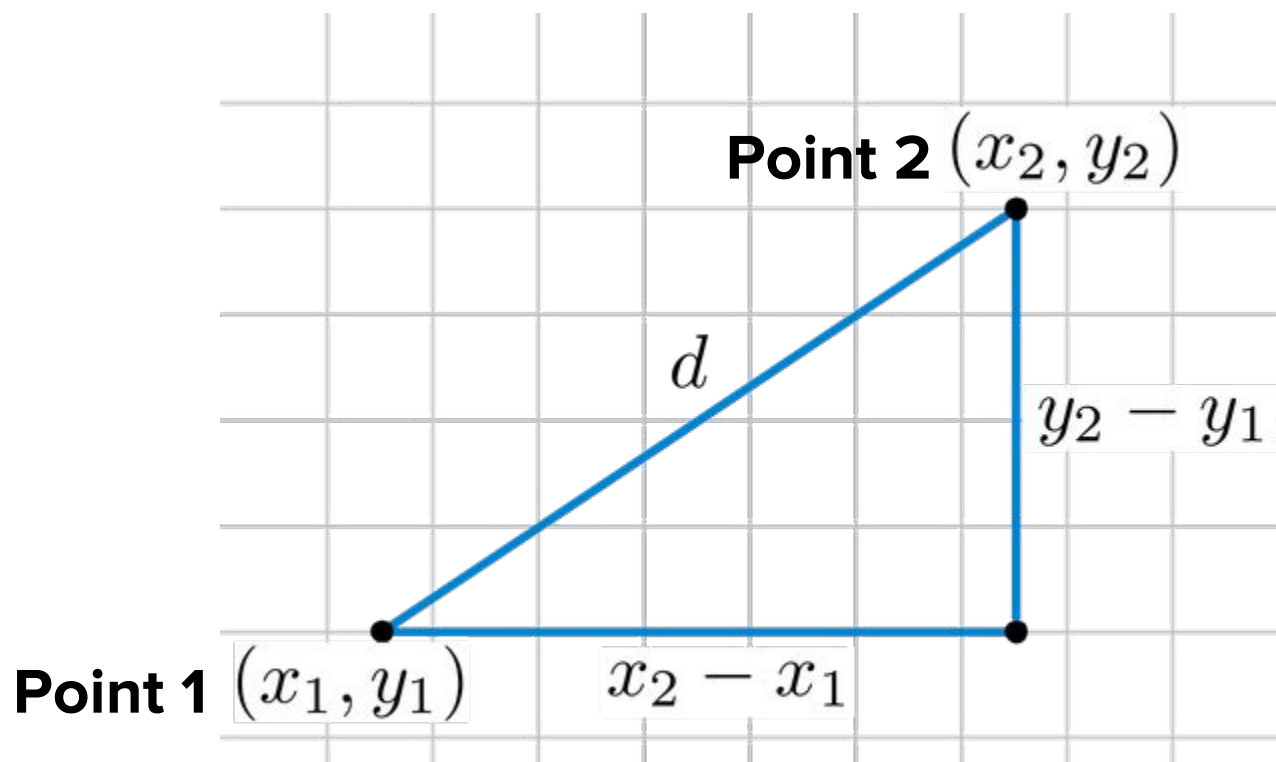


Euclidean Distance



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

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



Features scaling: Yes or Not?

Features scaling: Yes!

Linearity

Is it a linear model?

No!...and this is good.

Best K?

Small K - what happens?

Small K - Noise + variance

Big K - what happens?

Big K - too “general”

Best K? Tuning!

The model

The model:

Linear regression = one formula

KNN = ???

The model:

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 number

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

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

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

1. Classification: predicting a number
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 number
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