AIBridge

Lecture 6

*Cla*ssification!

Classification!

quick review
wine dataset

- Fixed acidity
- Volatile acidity
- Citric acid
- Residual sugar
- Chlorides
- Free sulfur dioxide
- Total sulfur dioxide
- Density
- pH
- Sulphates
- Alcohol

→ Model →

White = 0

Red = 1

a class

- categorical label outputs **are named "classes"**

*Cl*assification!

quick review

wine dataset

that's a lot
of features!

- Fixed acidity
- Volatile acidity
- Citric acid
- Residual sugar
- Chlorides
- Free sulfur dioxide
- Total sulfur dioxide
- Density
- pH
- Sulphates
- Alcohol

→ Model →

a class

White = 0

Red = 1

- categorical label outputs are named "classes"

- Fixed acidity
- Volatile acidity
- Citric acid
- Residual sugar
- Chlorides
- Free sulfur dioxide
- Total sulfur dioxide
- Density
- pH
- Sulphates
- Alcohol

that's a lot of features!

can they really be linear?
even in high dimensions?

- as feature counts increase, and on complex data, linear type model may not be the best model

as feature counts increase, and on complex data,
linear type models may not be the best model

we need **a more complex model**

Decision Trees

Decision Trees

Decision Trees

Can I afford it?

Decision Trees

Can I afford it?

Is it comfortable?

Decision Trees

Can I afford it?

Is it comfortable?

Is it fashionable?

Decision Trees

Can I afford it?

Is it comfortable?

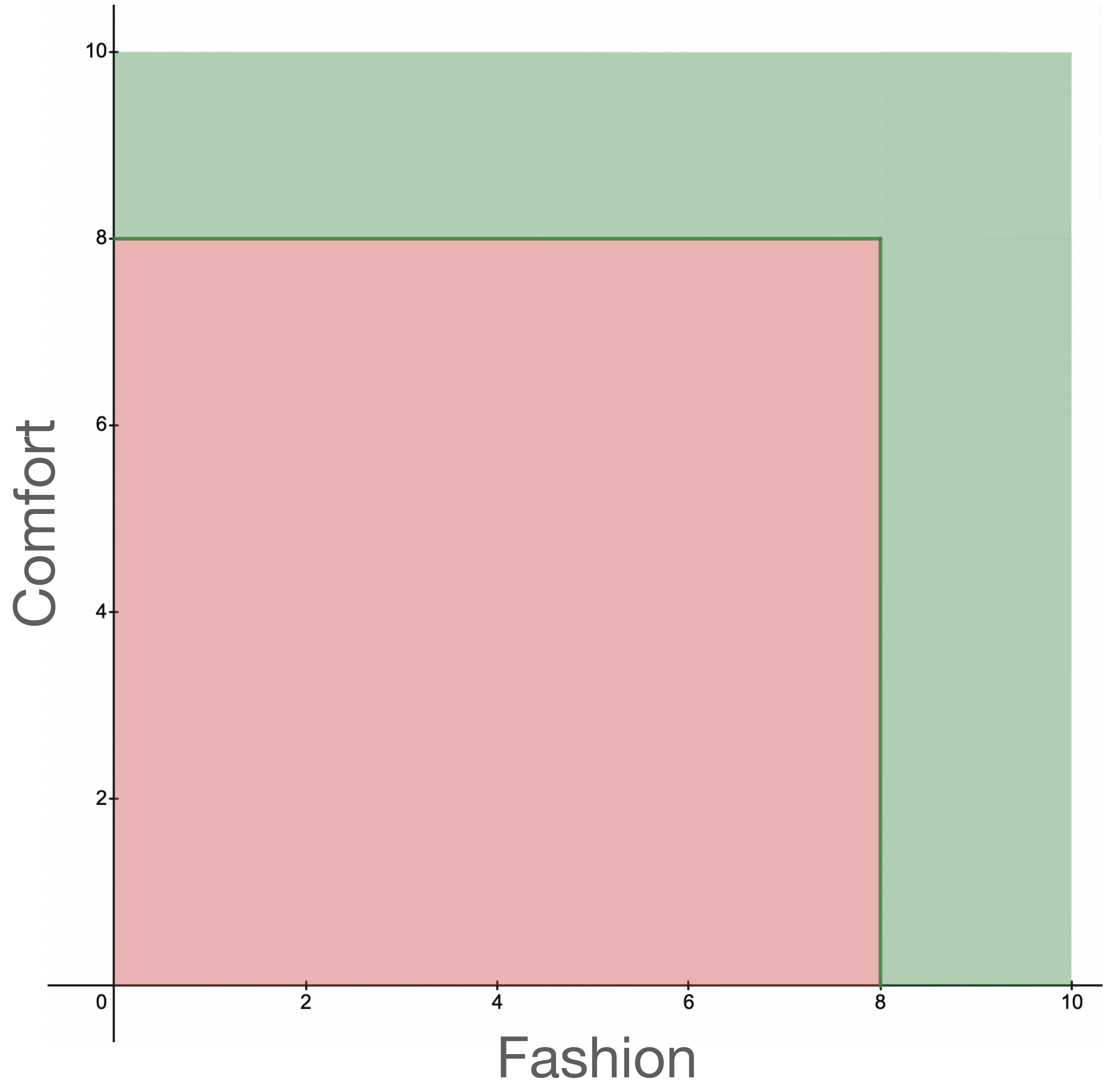
Is it fashionable?

Decision Trees

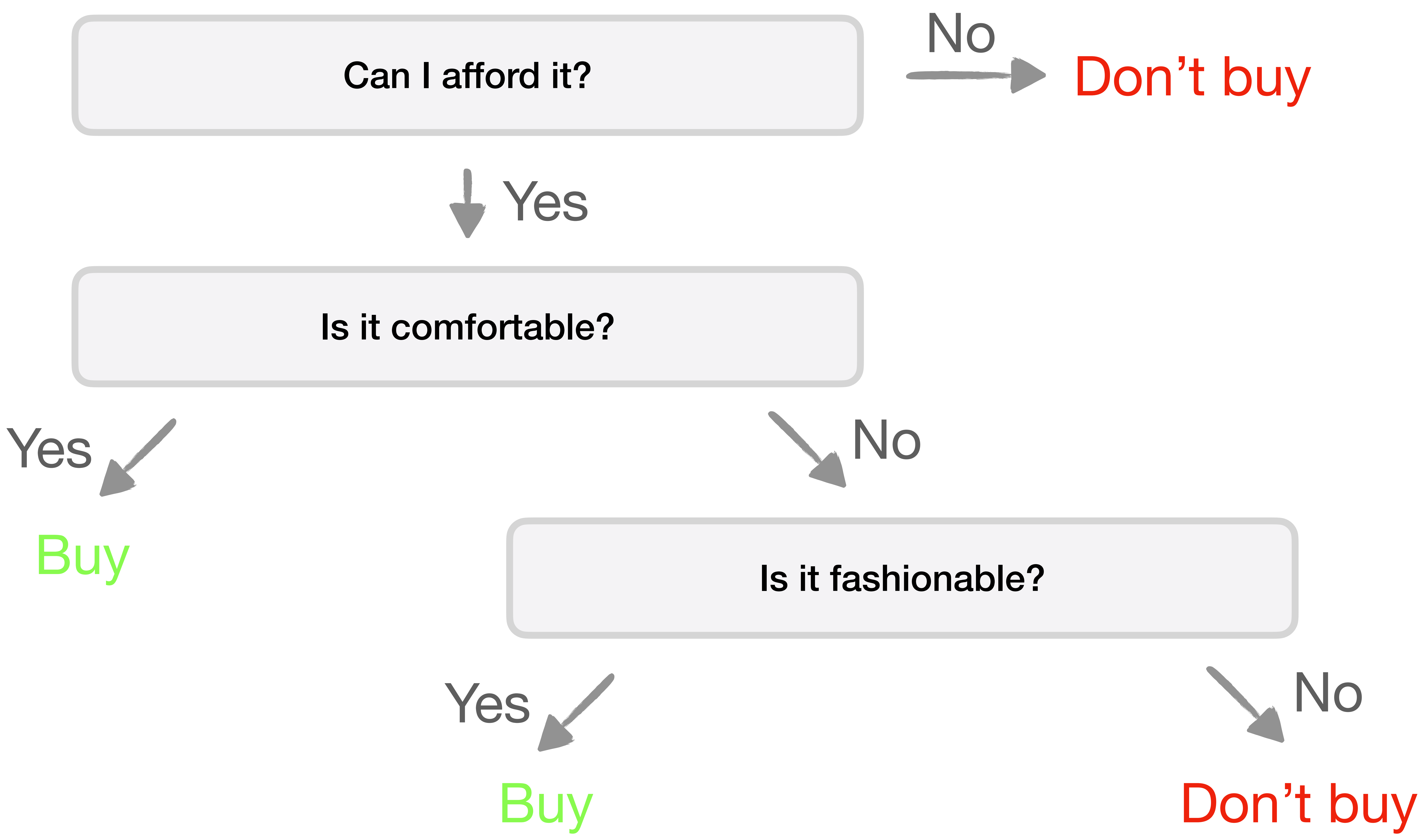
Can I afford it?

Is it comfortable?

Is it fashionable?



Decision Trees



Decision Trees

that seems awfully hard-coded!

- flowcharts of decisions can create an explainable and repeatable graph of predictions

Decision Trees

Price	Comfort	Fashion	Purchased?
$70	4	6	No
$120	5	8	No
$20	4	4	No
$60	1	8	Yes
$60	6	3	No
$80	8	8	Yes

Decision Trees

Purchased?

No

No

No

Yes

No

Yes

Decision Trees

No	Yes
No	No
No	Yes

Decision Trees

No	Yes
No	No
No	Yes

Decision Trees

Which one is a better split?

No	
No	No
No	Yes
Yes	

Mostly no

No	Yes
No	No
No	Yes

All no

Mostly

(Gini impurity)

Gini impurity

	Purchased?	Purchased?	
	No	No	
	No	No	
0.38	No	No	0
	Yes		
		Yes	
0.5	No	No	0.44
	Yes	Yes	

- as a group becomes more **homogeneous**, its **Gini Impurity** decreases.

Gini impurity

	Purchased?
	No
	No
0.38	No
	Yes
	No
0.5	Yes

Purchased?	
No	
No	
No	0
Yes	
No	0.44
Yes	

- as a group becomes more **homogeneous**, its **Gini Impurity** decreases.
- perfect groups => 0 **Gini Impurity** => 100% predictions

Portion of that one class in group

Portion of not that one class in the group

$$G = \sum_{i=1}^{C} P(i) \cdot (1 - P(i)) \blacksquare$$

Add them up for all groups

- **Gini impurity** measures the homogeneity in a group

Decision Trees

Purchased?

No

No

0

No

Yes

No

Yes

0.5

0.5

Decision Trees

Purchased?	
No	
No	0.38
No	
Yes	
No	0.5
Yes	
	0.88

Decision Trees

we gotta do better
than this, right?

Purchased?

No	
No	0
No	
Yes	
No	0.44
Yes	
	0.44

just split again!

Purchased?	
No	
No	0
No	
Yes	
No	0.44
Yes	
	0.44

Decision Trees

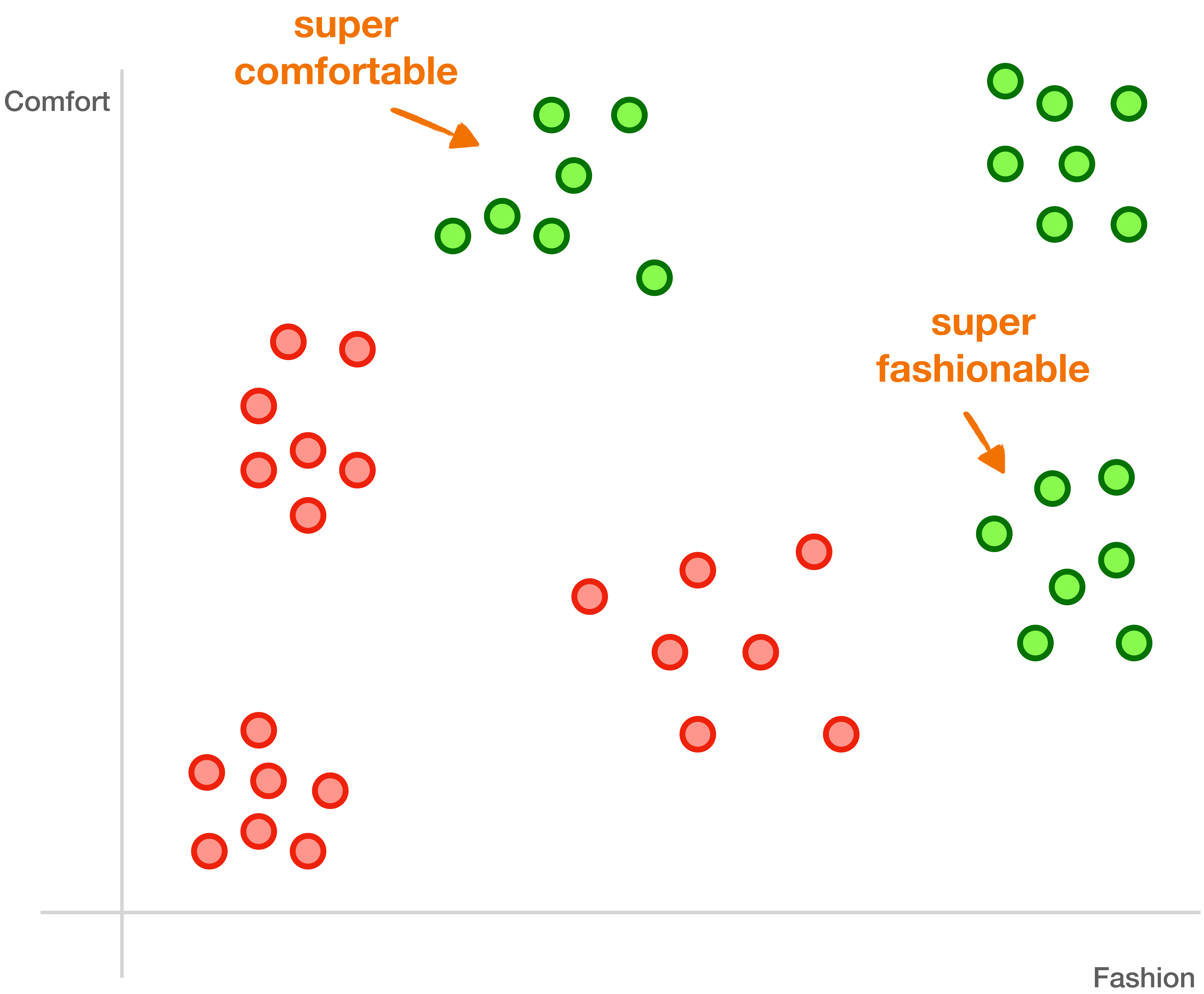
1. Make splits and calculate Gini impurity
2. Select the split with the lowest Gini impurity
3. If unhappy, **just split again!**
4. Repeat 1-3 as much as needed

a hyperparameter

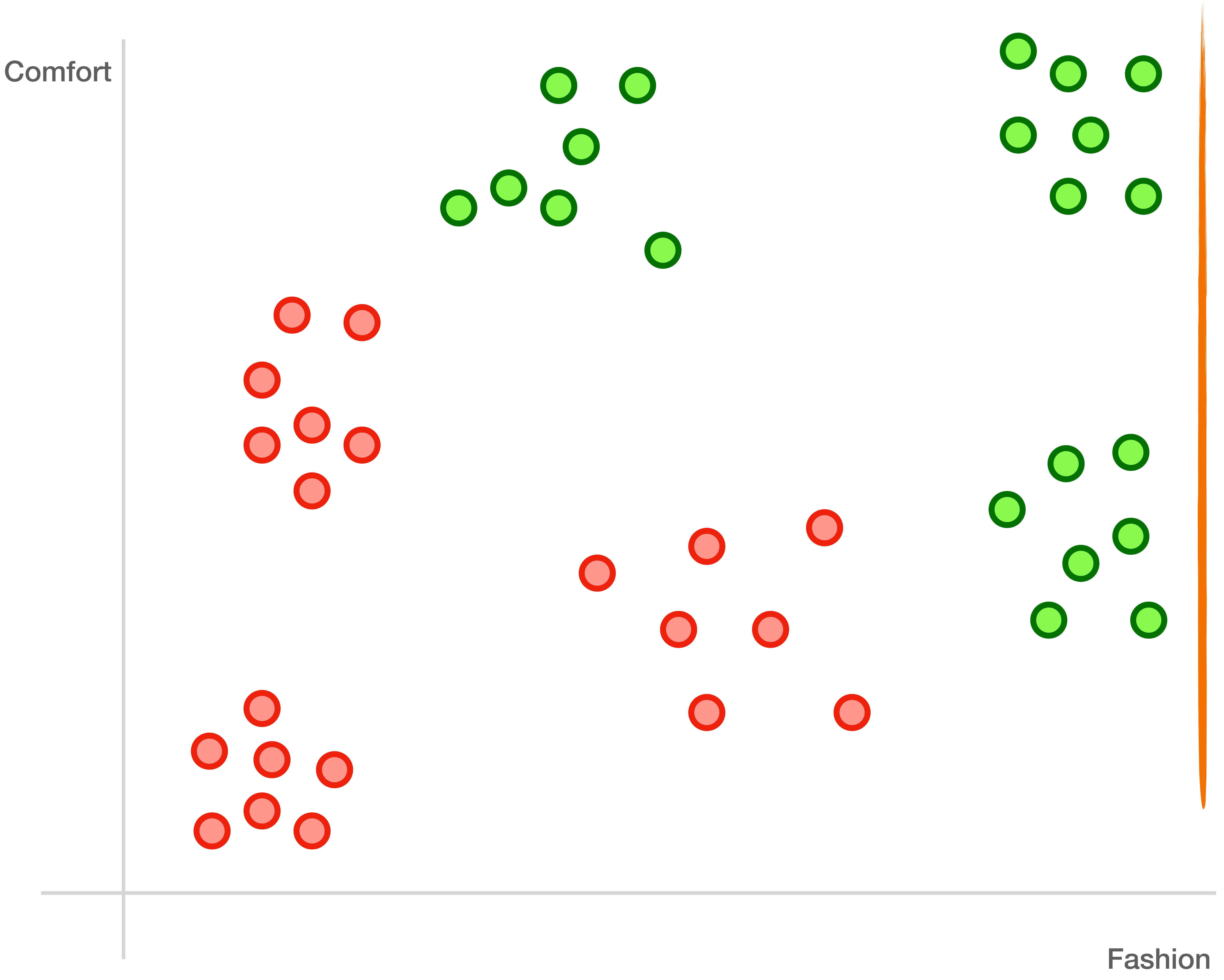
■

“split”

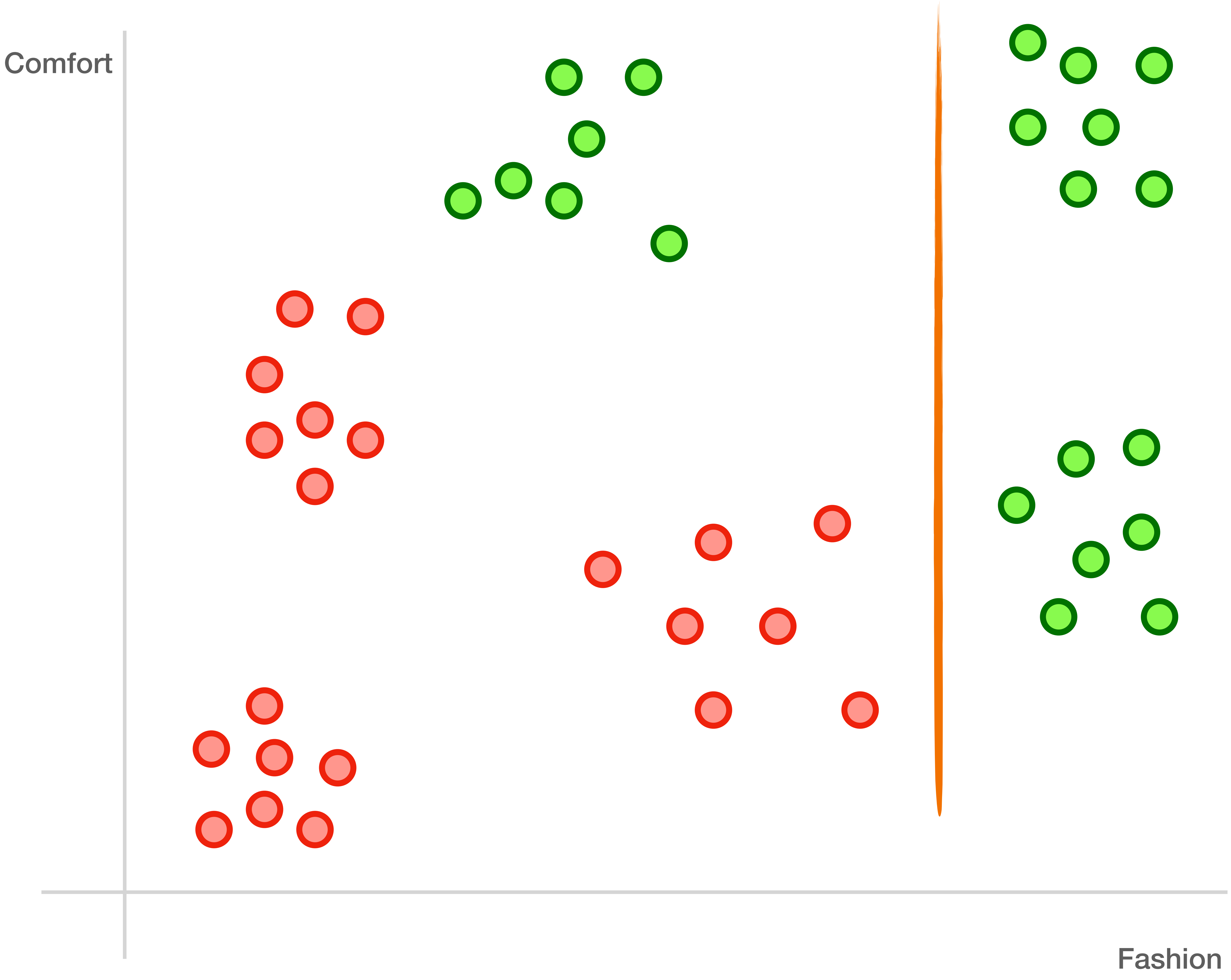
“split”



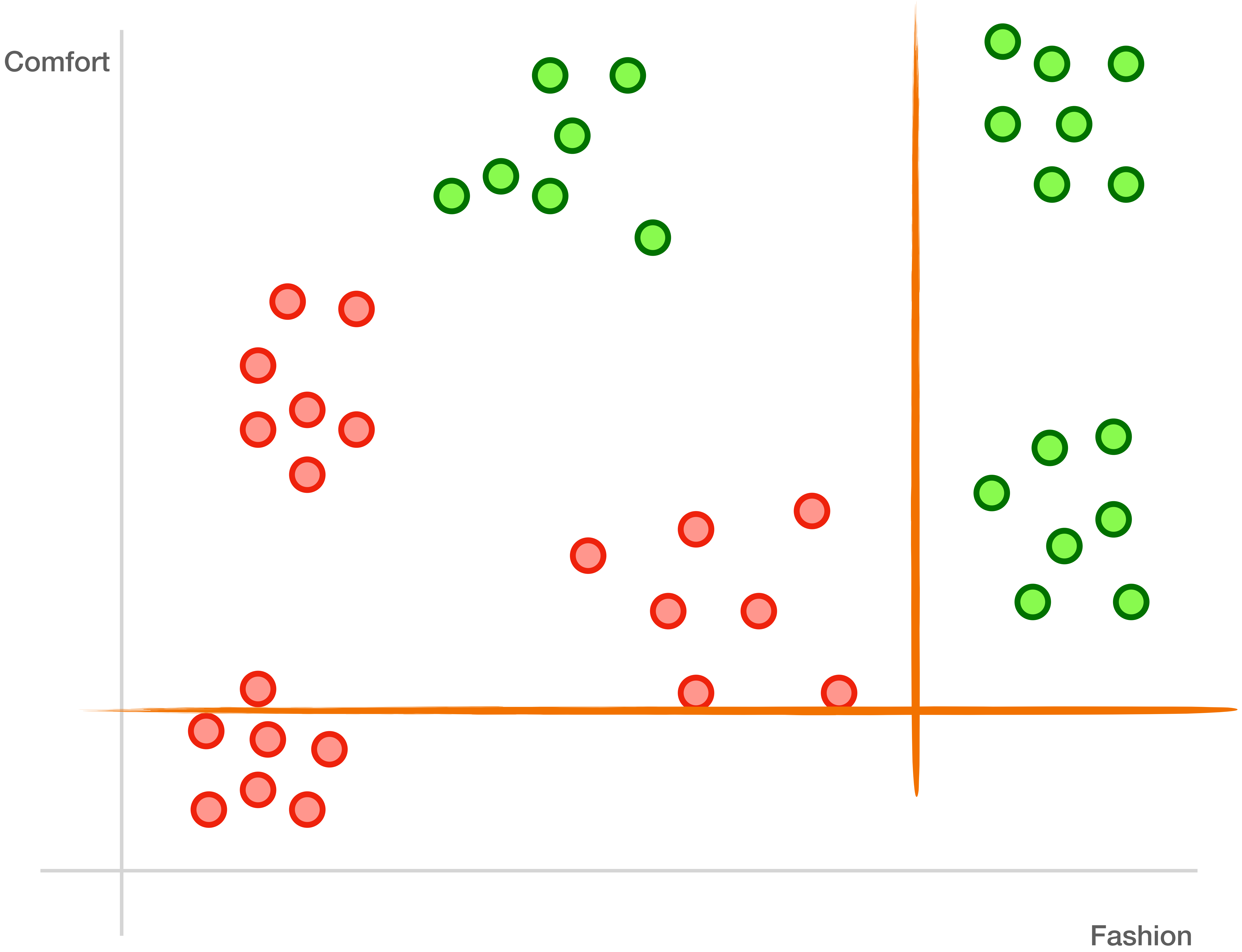
"split"



“split”

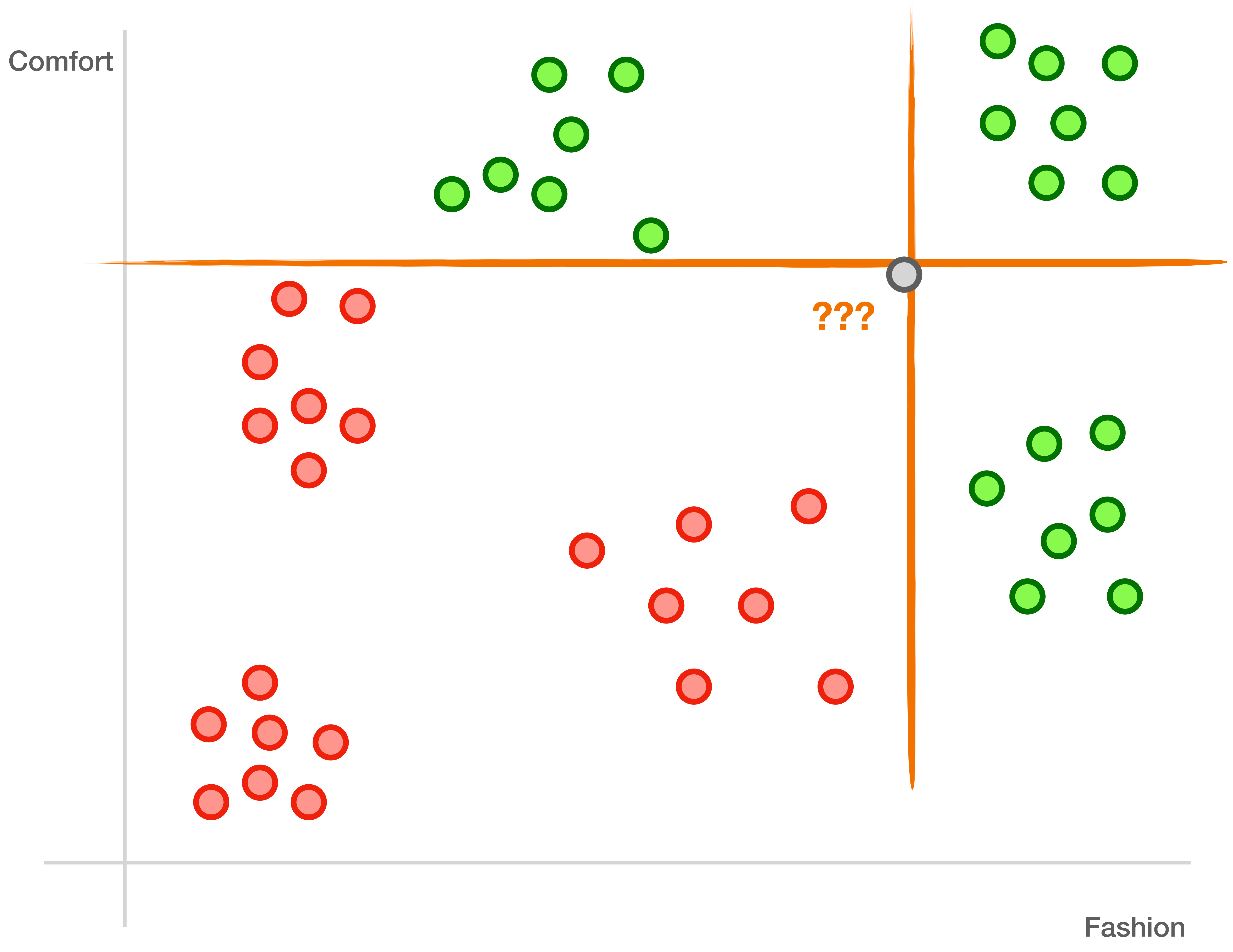


“split”



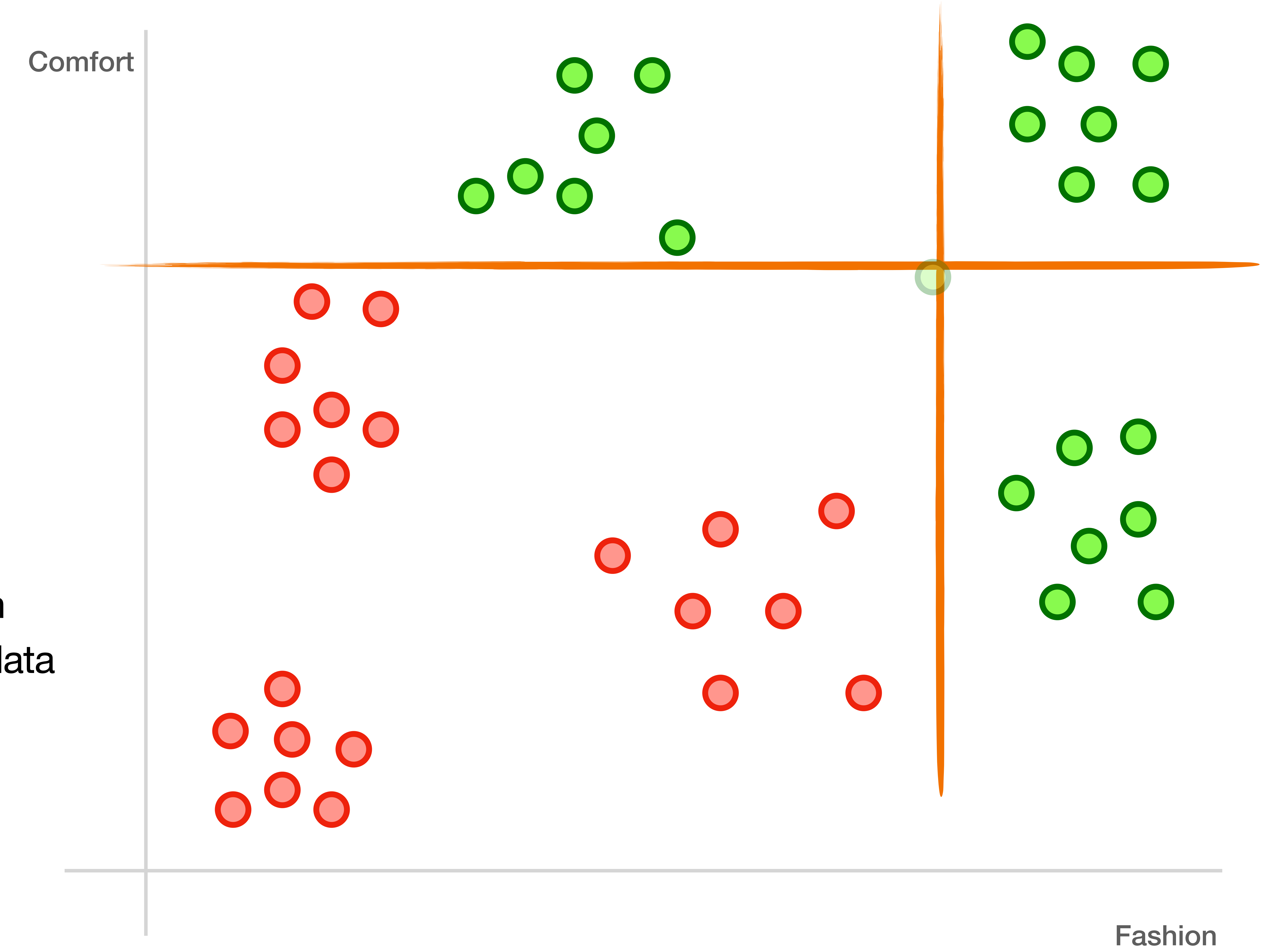
this is fine,
but…

“split”



“split”

- the **binary splits** in decision trees often don’t do well in complex, multivariate data

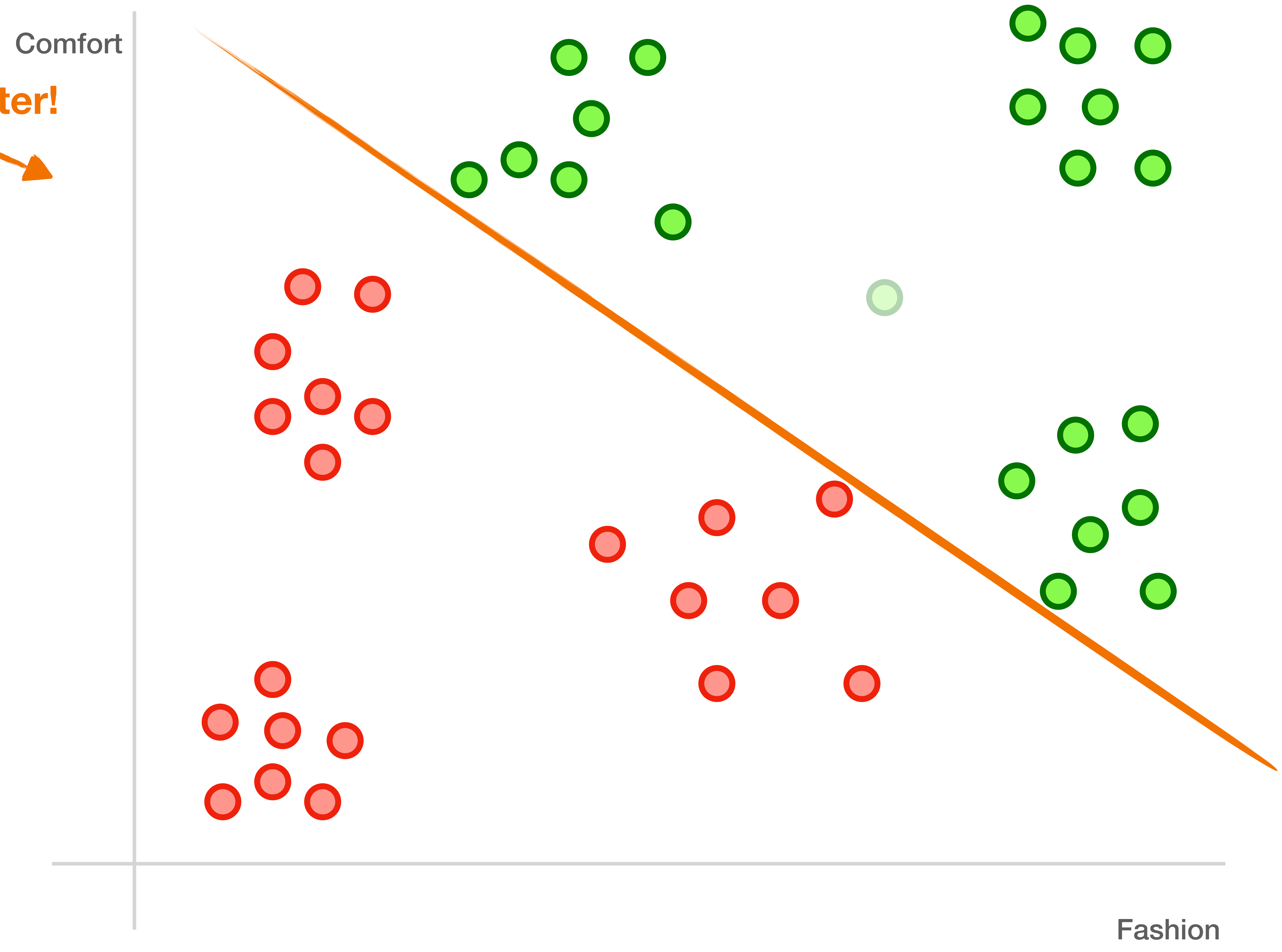


we need a more complex model

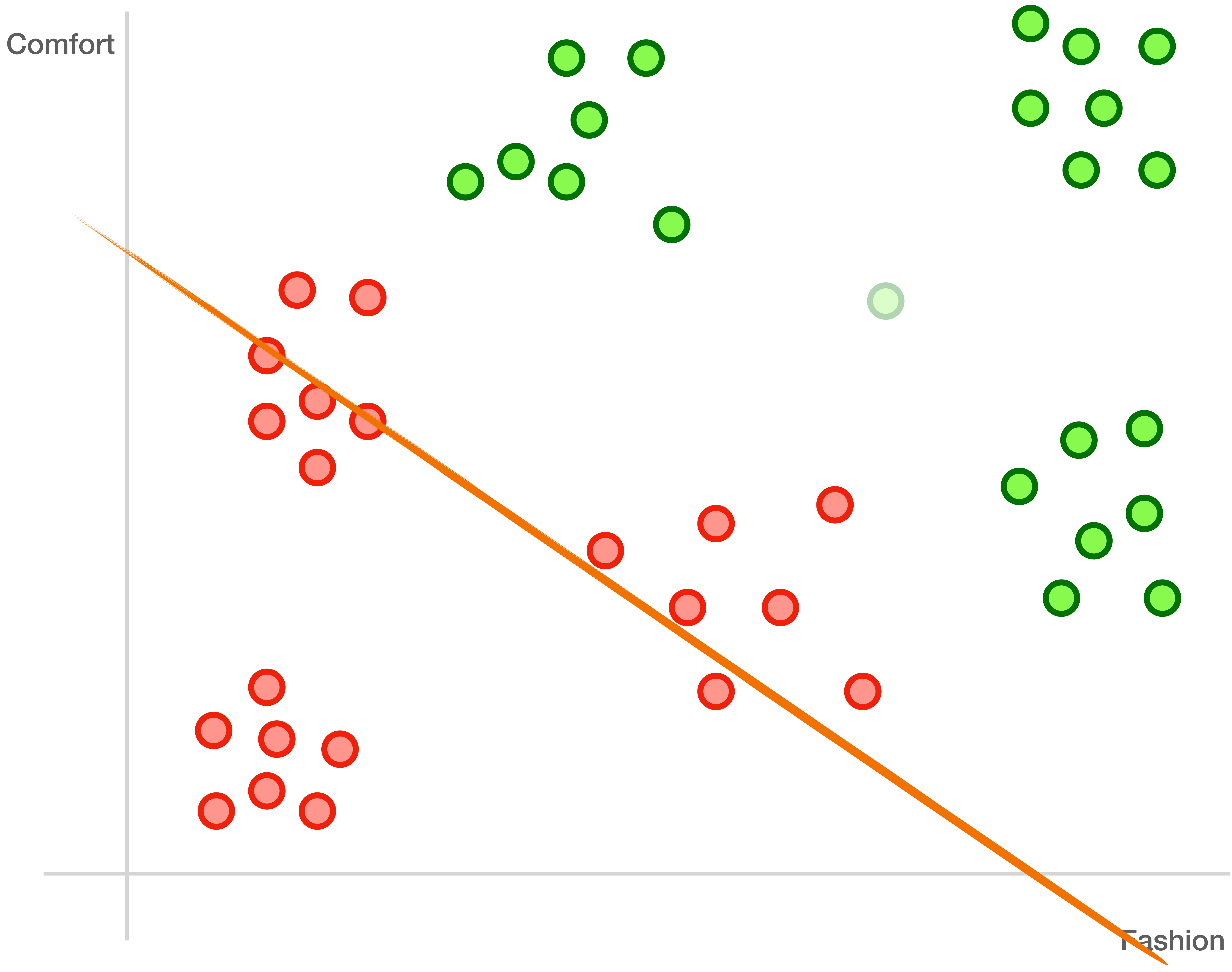
Support vector machines!

Support vector machines!

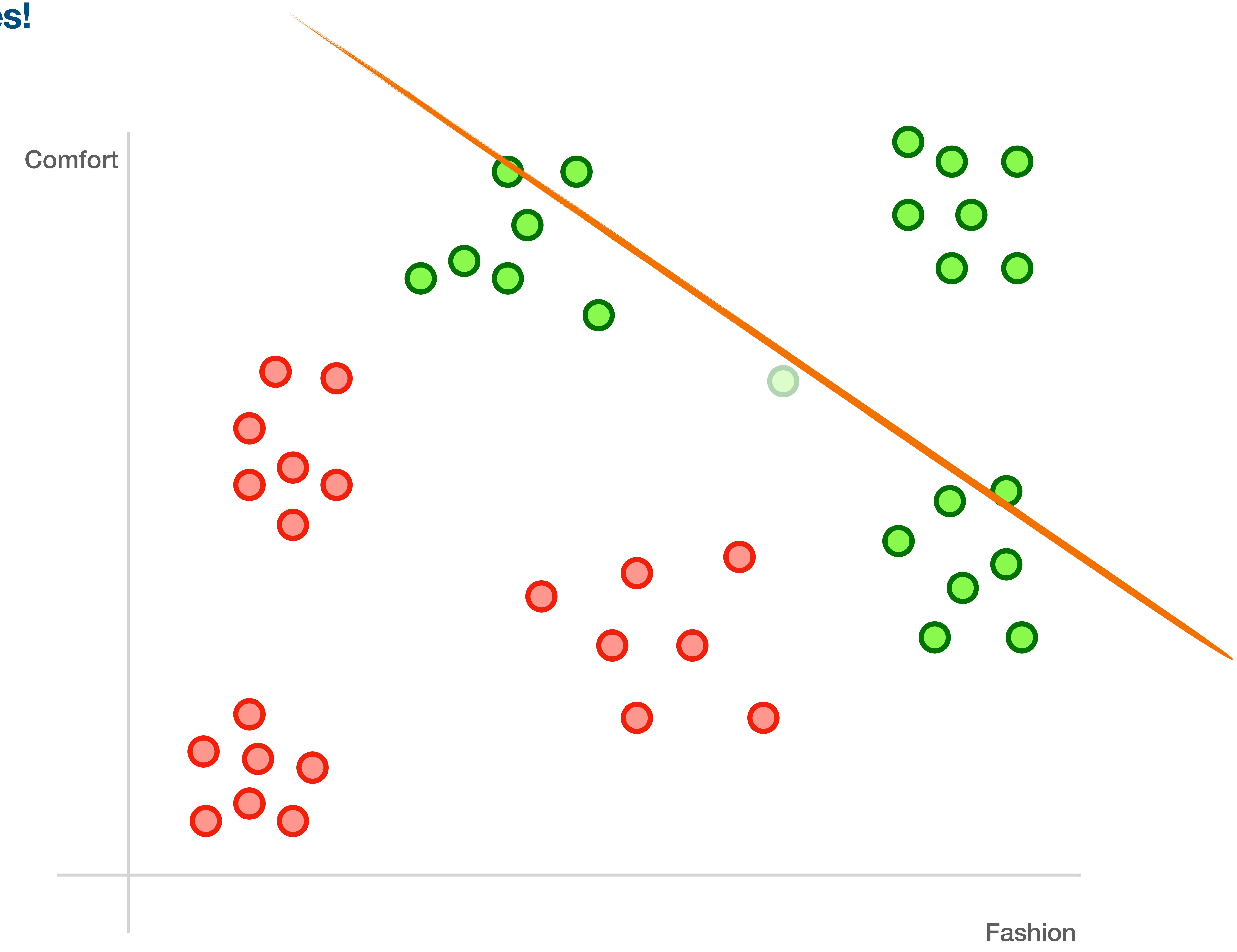
much better!



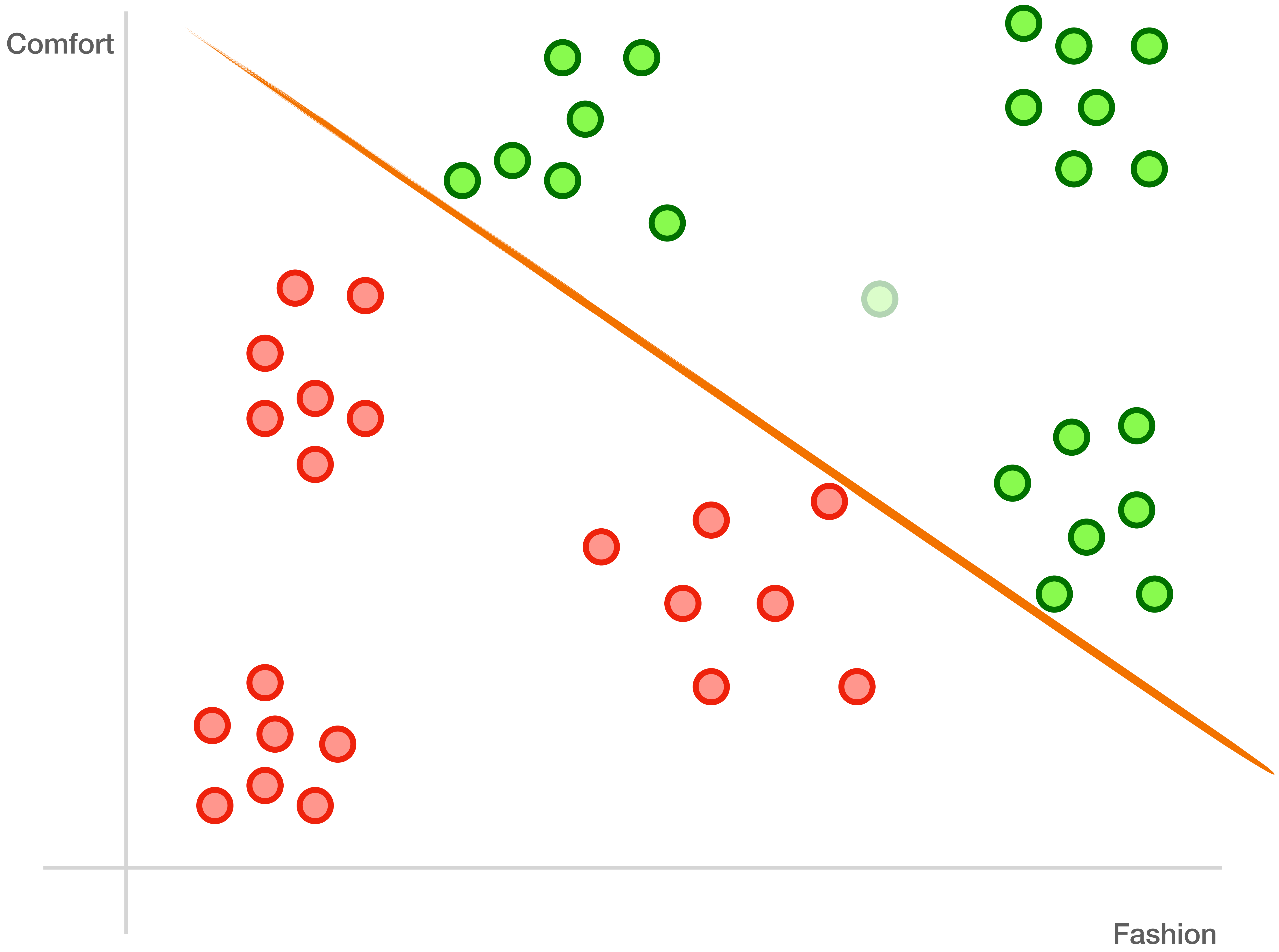
Support vector machines!

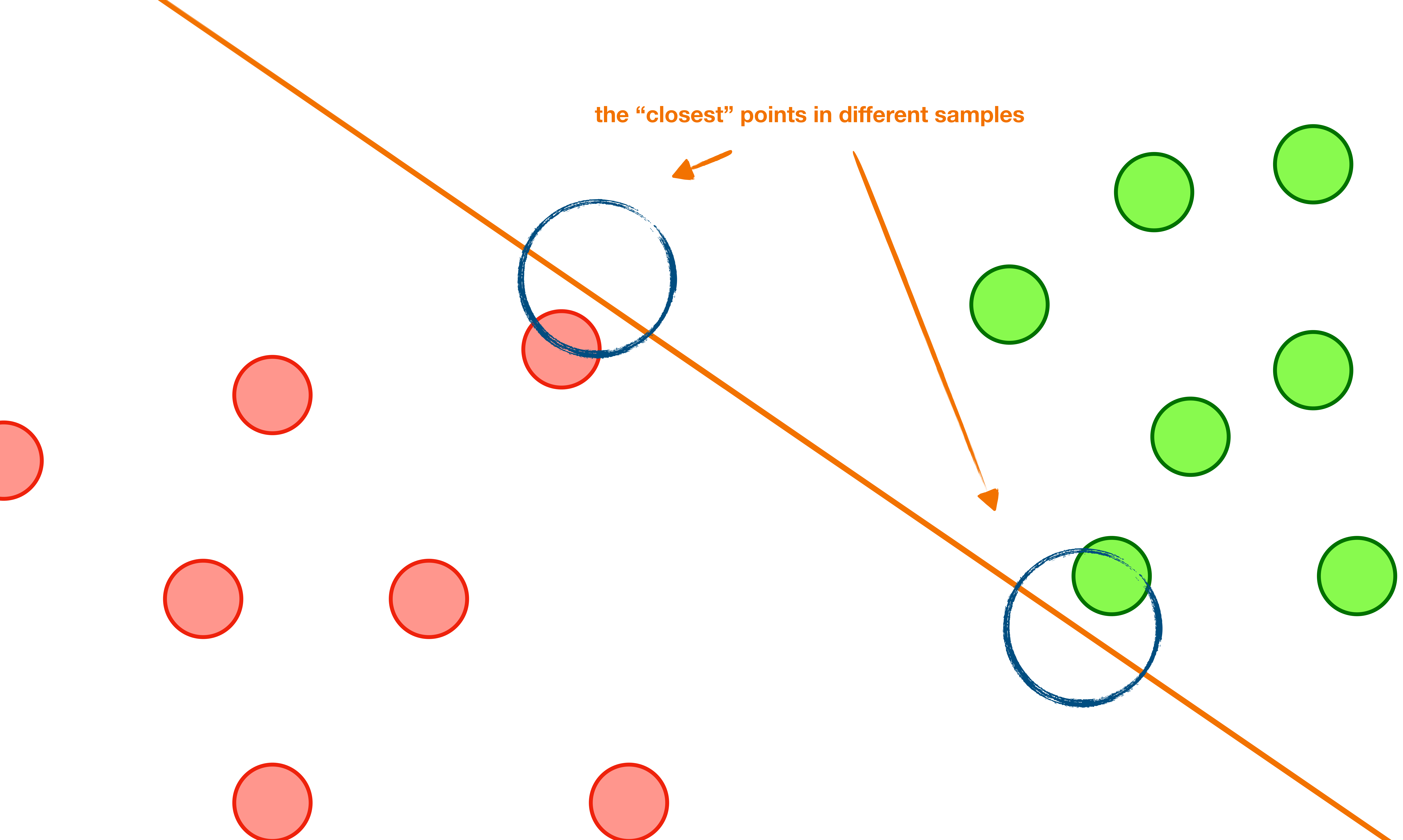


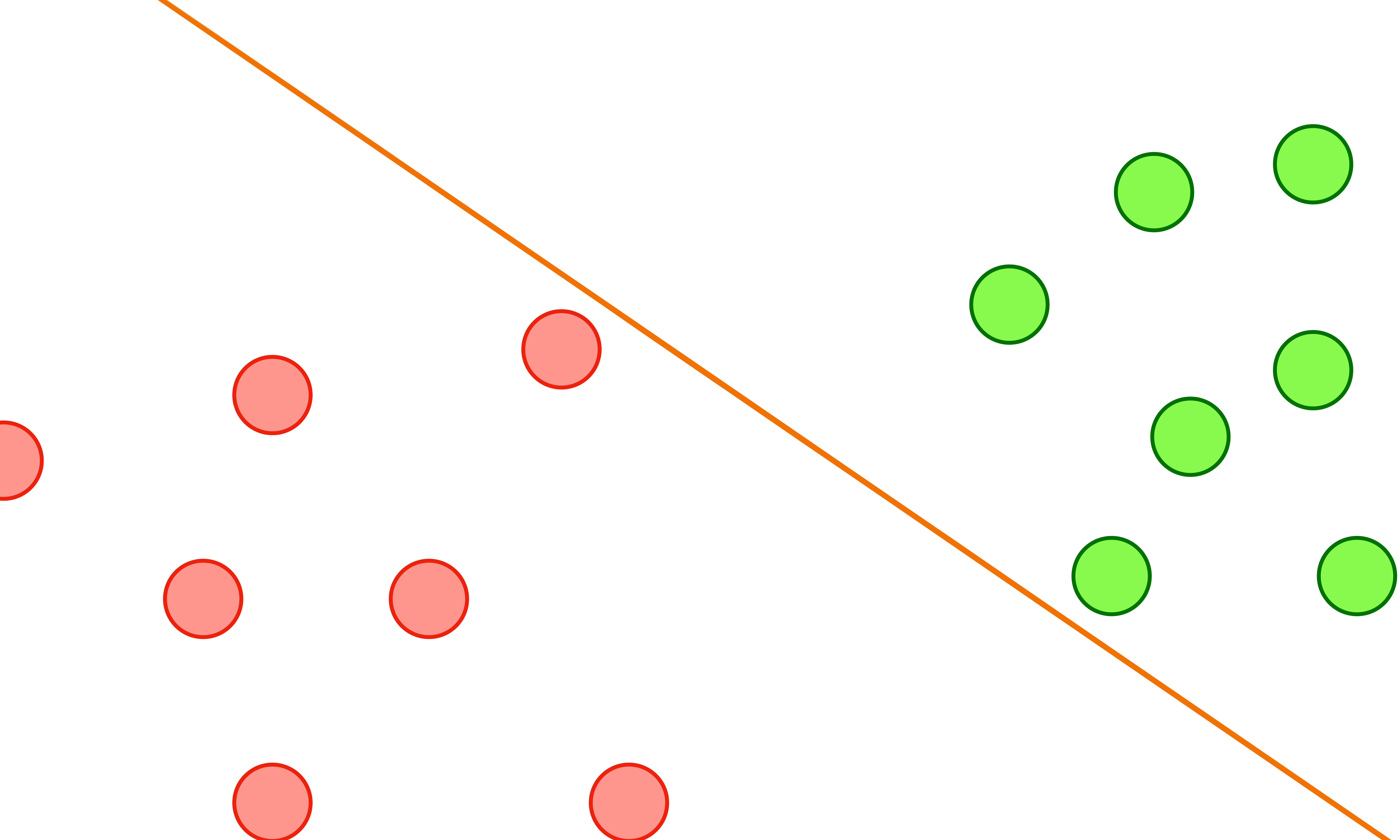
Support vector machines!

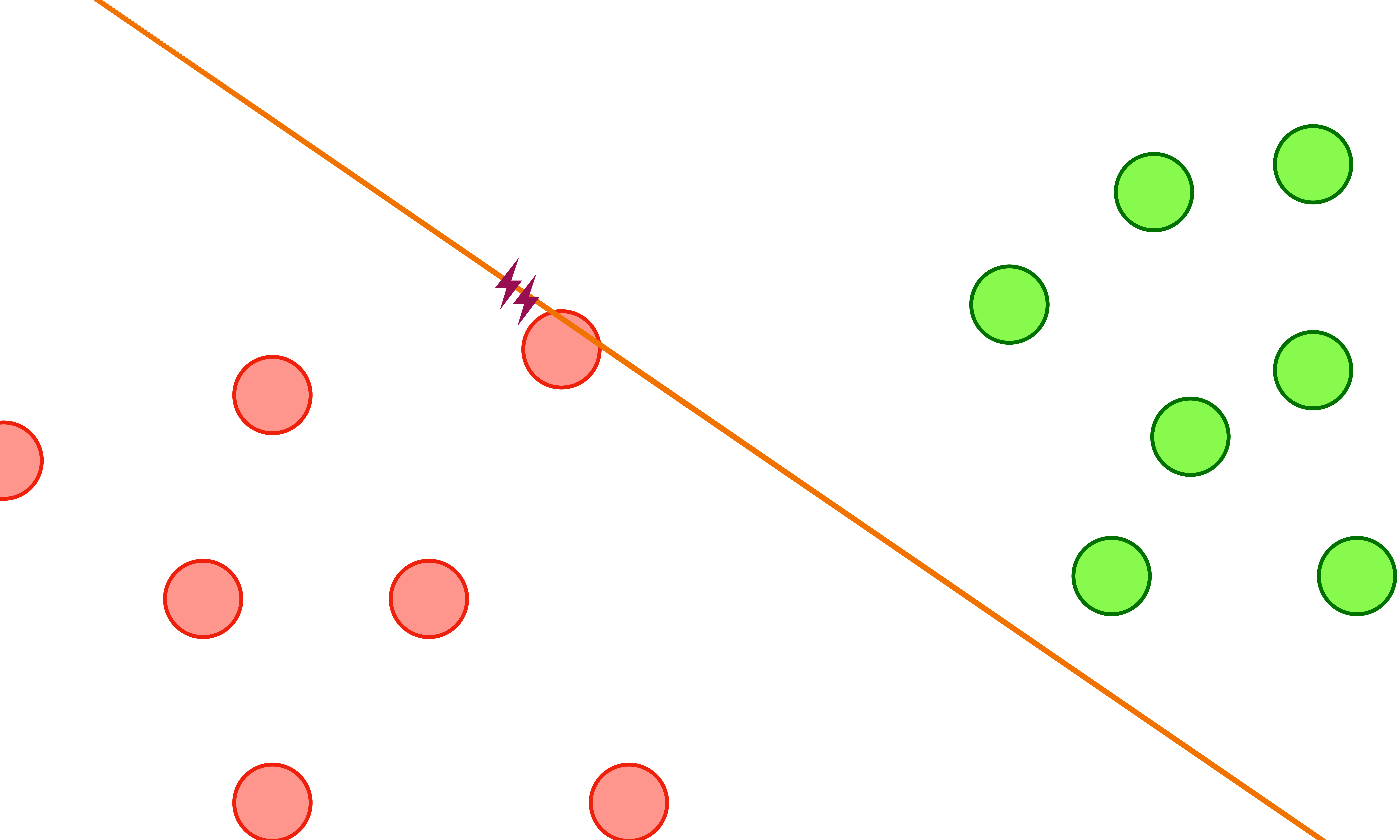


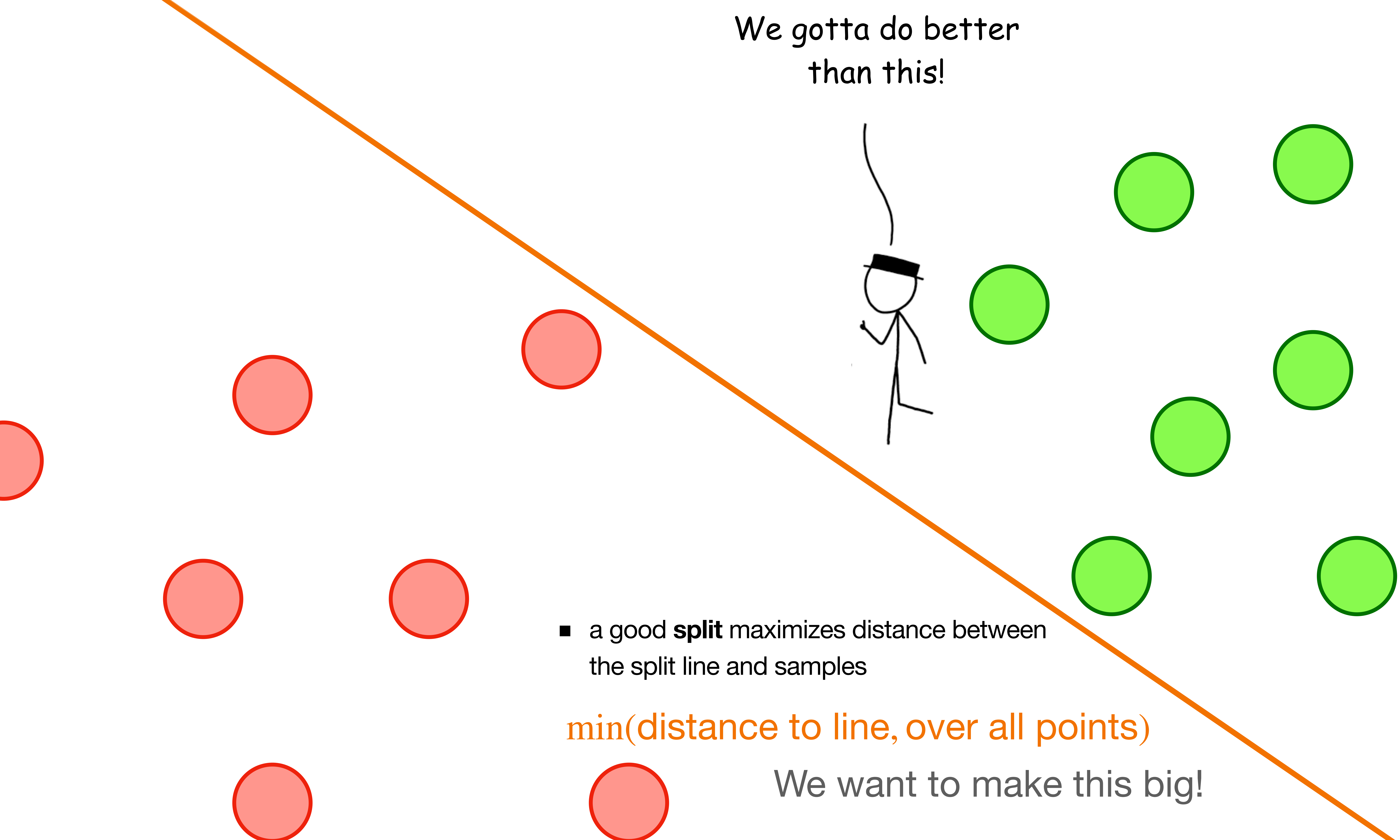
Support vector machines!




the “closest” points in different samples







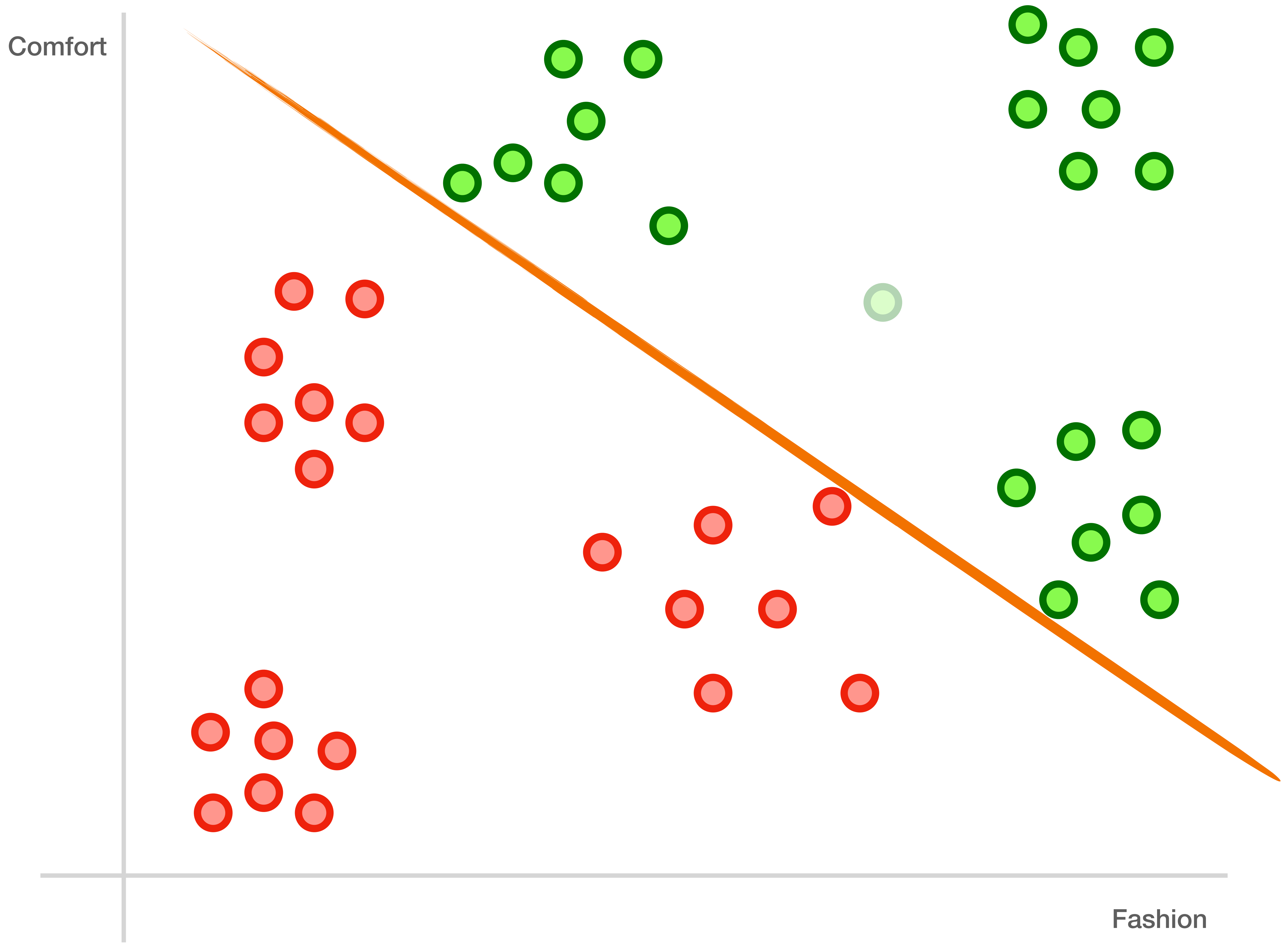
- a good **split** maximizes distance between the split line and samples

$\min($distance to line, over all points$)$

We want to make this big!

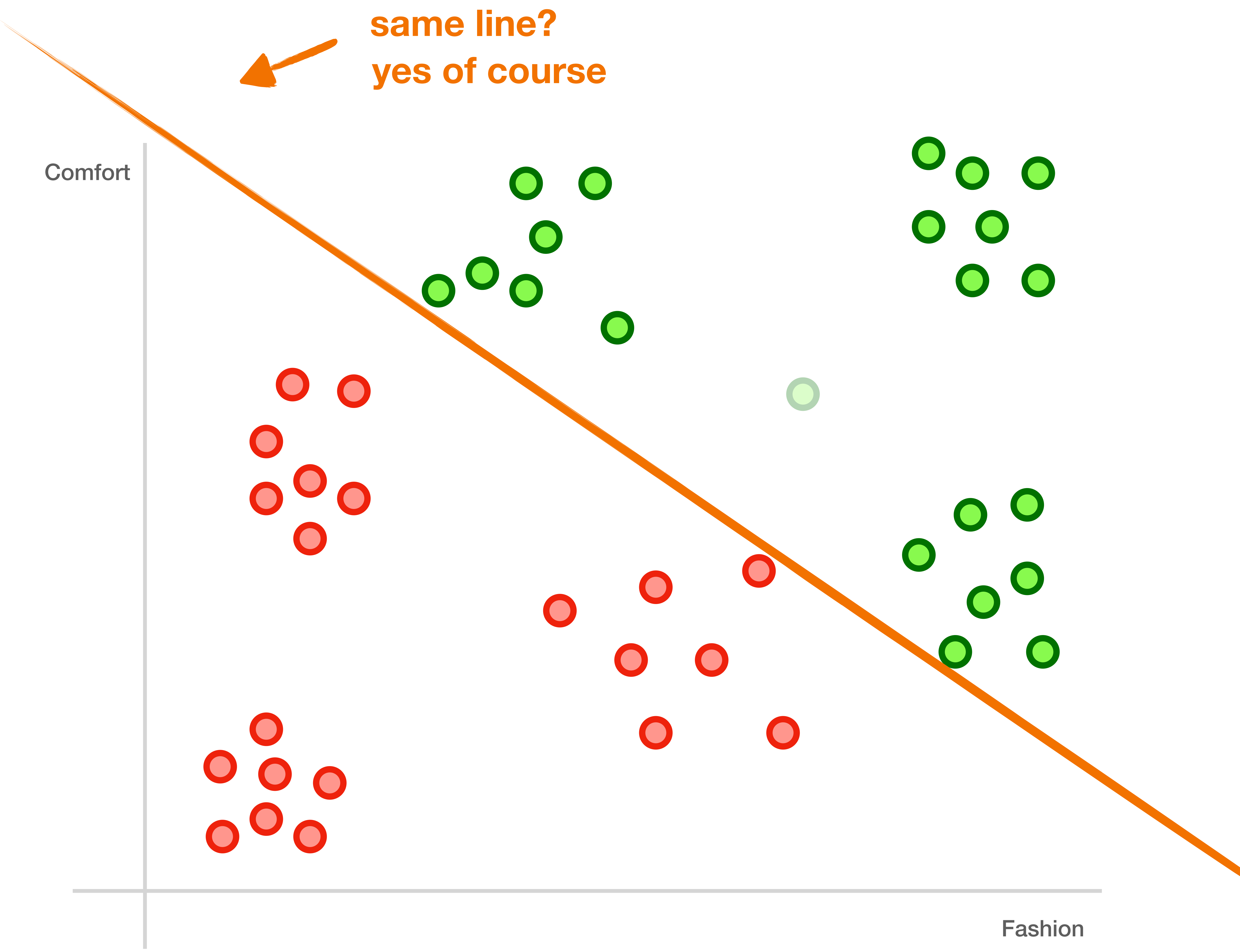
Support vector machines!

$\max(\text{distance to line, over all points})$



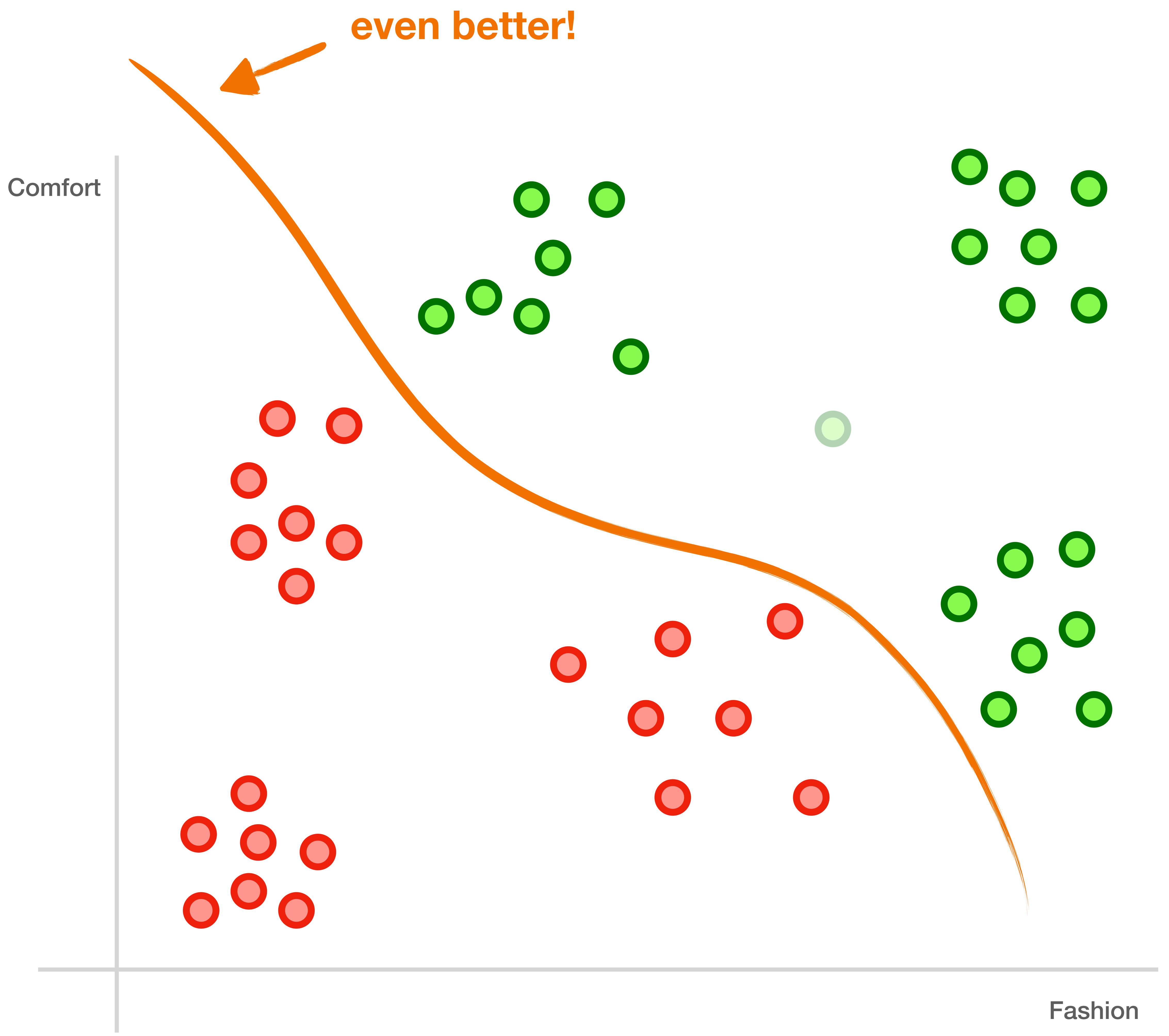
Support vector machines!

$\max(\text{distance to line, over all points})$



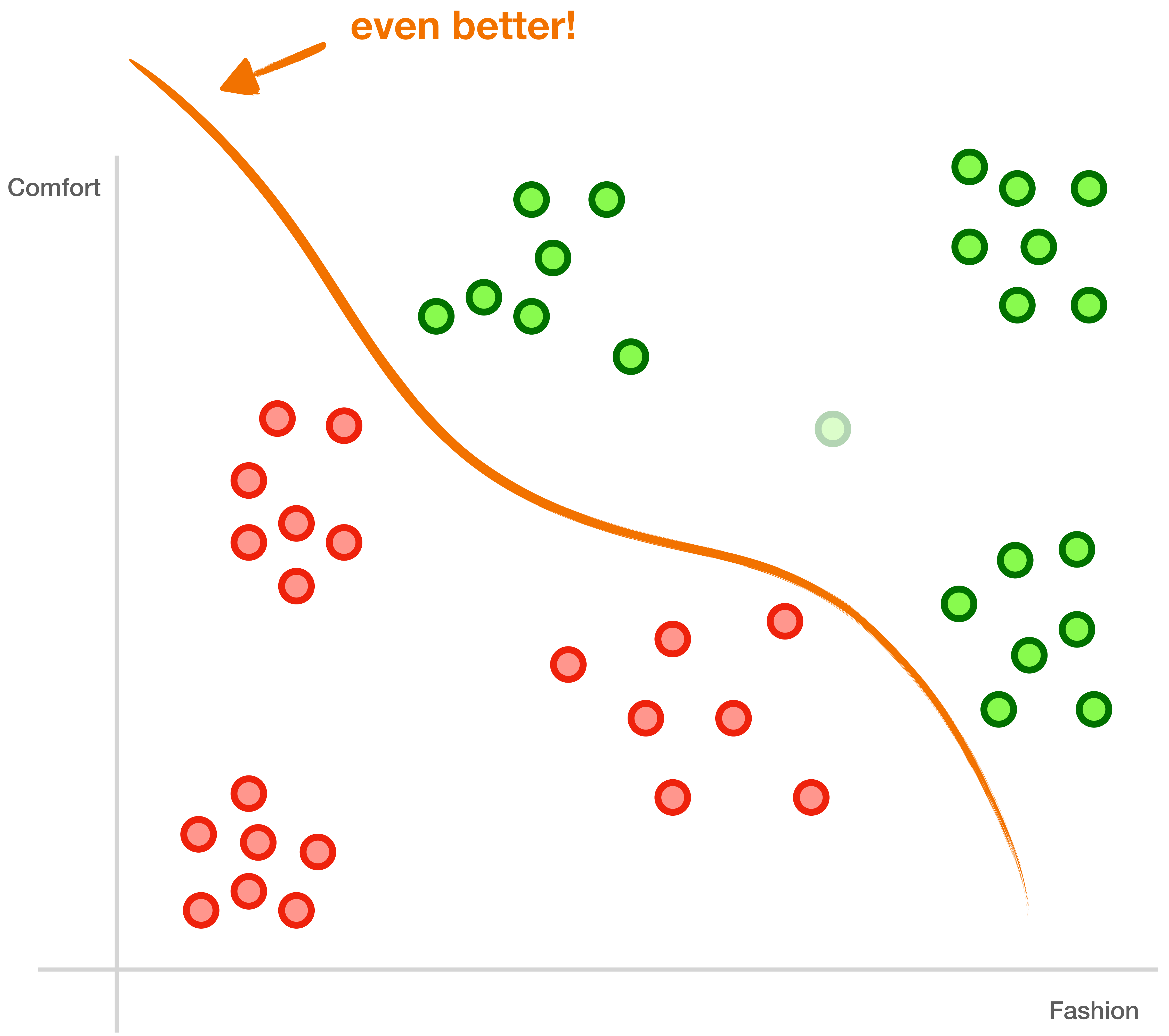
Support vector machines!

$\max(\text{distance to line, over all points})$



Support vector machines!

$\max(\text{distance to line, over all points})$



- **support-vector machines** are classifiers that divide data by class, aiming to create a **margin** that's as wide as possible.
- They use non-linear functions

BACK ← TO THE PROBABILITY

Internal Memo:

146 Hagley Road, Birmingham
Birmingham B3 3PJ

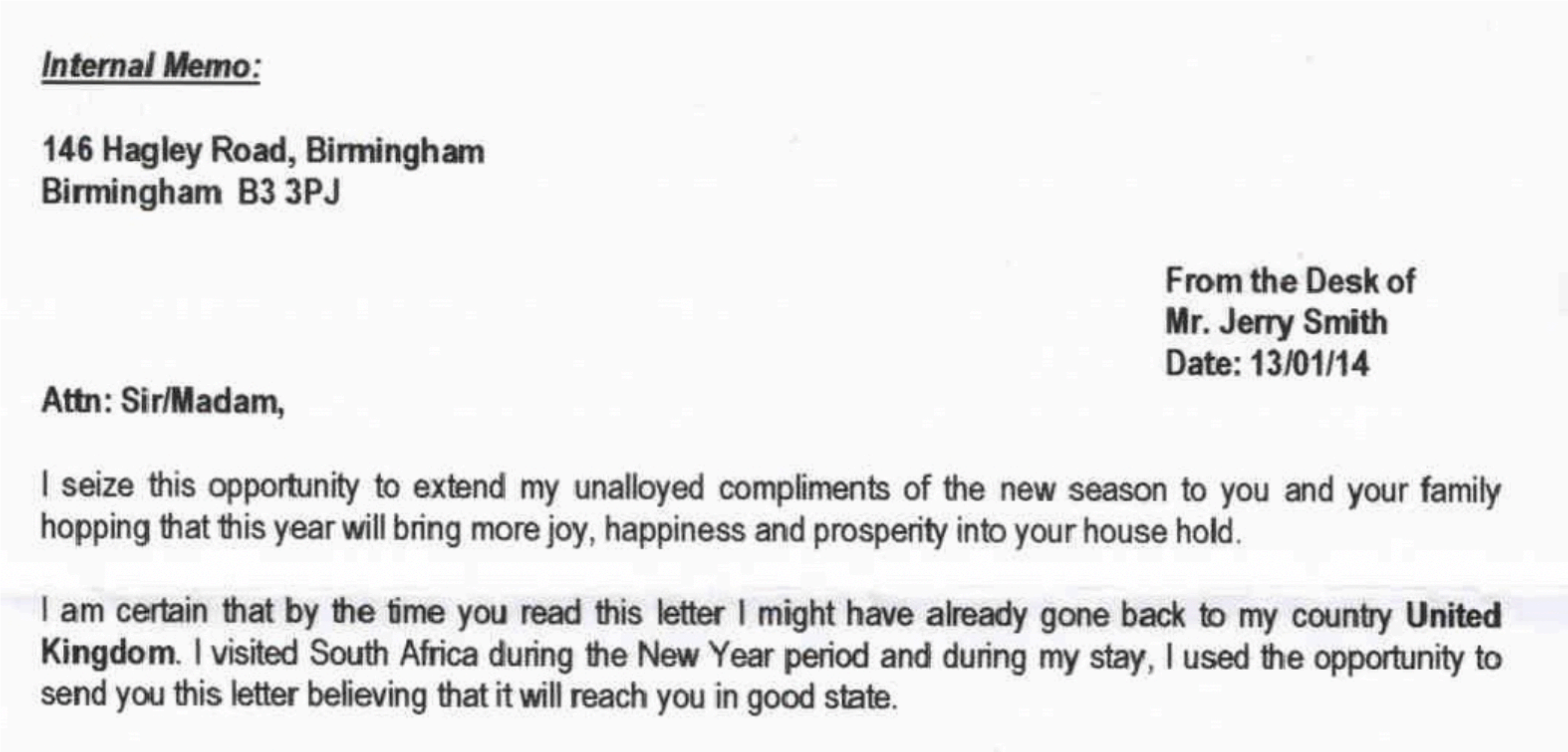
From the Desk of
Mr. Jerry Smith
Date: 13/01/14

Attn: Sir/Madam,

I seize this opportunity to extend my unalloyed compliments of the new season to you and your family hopping that this year will bring more joy, happiness and prosperity into your house hold.

I am certain that by the time you read this letter I might have already gone back to my country **United Kingdom**. I visited South Africa during the New Year period and during my stay, I used the opportunity to send you this letter believing that it will reach you in good state.

PROBABILITY


Internal Memo:

146 Hagley Road, Birmingham
Birmingham B3 3PJ

From the Desk of
Mr. Jerry Smith
Date: 13/01/14

Attn: Sir/Madam,

I seize this opportunity to extend my unalloyed compliments of the new season to you and your family hopping that this year will bring more joy, happiness and prosperity into your house hold.

I am certain that by the time you read this letter I might have already gone back to my country **United Kingdom**. I visited South Africa during the New Year period and during my stay, I used the opportunity to send you this letter believing that it will reach you in good state.

"unalloyed complements" → Spam

"$100,000 dollars" → Spam

"relative dying of cancer" → Spam

PROBABILITY

IF we have this		THEN we have this
“unalloyed complements”	→	Spam
“$100,000 dollars”	→	Spam
“relative dying of cancer”	→	Spam

IF we have this THEN we have this

IF we have this THEN we have this

$$A \mid B$$

IF we have this THEN we have this

$$A \mid B$$

- Is Spam
- “Nigerian Prince”

IF we have this THEN we have this

$$\textit{spam} \mid \textit{nigerian prince}$$

IF we have this THEN we have this

$$P(\mathit{spam} \mid \mathit{nigerian\ prince})$$

high?	Nigerian prince	→	spam likely
low?	Nigerian prince	→	not spam

- **conditional probabilities** can be used as a classifier!

$$P(\textit{spam} \mid \textit{nigerian prince}) = \frac{P(\textit{spam})P(\textit{nigerian prince} \mid \textit{spam})}{P(\textit{nigerian prince})}$$

PROBABILITY

% of spam in dataset

% of spam in dataset that relates to Nigerian prince

$$P(spam \mid nigerian\ prince) = \frac{P(spam)P(nigerian\ prince \mid spam)}{P(nigerian\ prince)}$$

% of Nigerian prince in dataset

Naïve Bayes Classifier

$$P(spam \mid nigerian\ prince, offer) = \frac{P(spam)P(nigerian\ prince \mid spam)}{P(nigerian\ prince)} \frac{P(offer \mid spam)}{P(offer)}$$

- **conditional probabilities** can be used as a classifier!
- a classifier made this way, however, is "**naïve**" when extended to multiple features

multiplication for AND assumes independence!
"naïve"

Three classifiers! That's a lot.

Let's get to the *long lab!*