AIBridge

Lecture 6

Classification!

- 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”**

Classification!

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

- Linear models might not be the best in some cases

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?

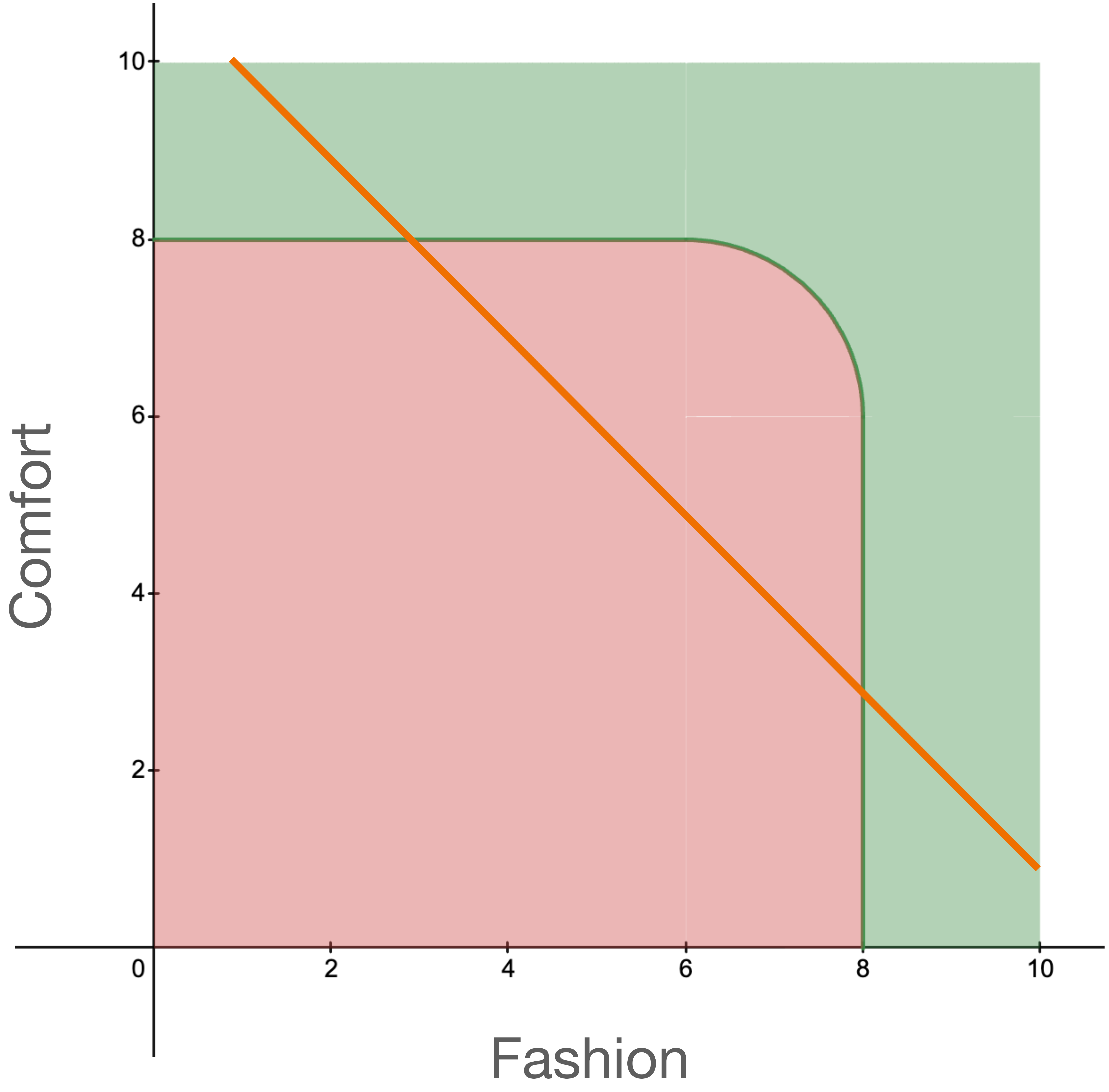
Is it fashionable?

Decision Trees

Can I afford it?

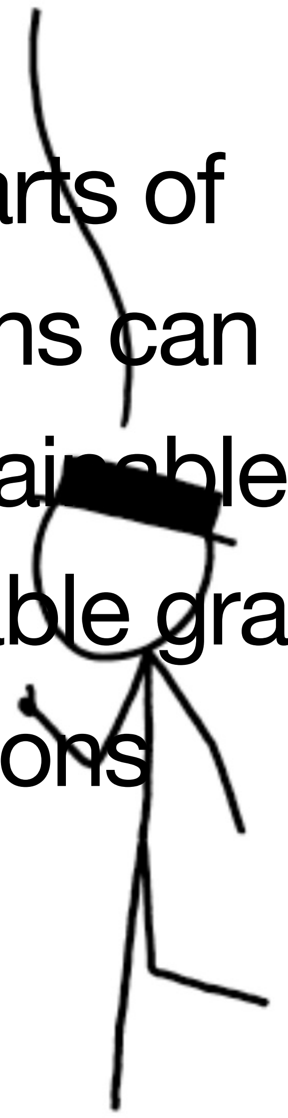
Is it comfortable?

Is it fashionable?



Decision Trees

Decision Trees



- 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?

(when trying to automate Decision Trees)

No	Yes
No	No
No	Yes

All no

No	
No	No
No	Yes
Yes	

Mostly no

Decision Trees

Purchased?

0

No
No
No

0.44

Yes
No
Yes

Purchased?

0.38

No
No
No
Yes

0.5

No
Yes

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

Decision Trees

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

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

Decision Trees

Fraction of that one class in group

Fraction of not that one class in the group

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

Add them up for all classes (in one side of the split)

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

Decision Trees

Purchased?	
No	0
No	
No	0.5
Yes	
No	
Yes	
	0.5

Decision Trees

Purchased?

No

No

No

Yes

0.38

No

Yes

0.5

0.88

Decision Trees

we gotta do better
than this, right?

Purchased?	
No	
No	0
No	
Yes	
No	0.44
Yes	
	0.44

Decision Trees

just split again!

Purchased?	
No	
No	0
No	
Yes	
No	0.44
Yes	
	0.44

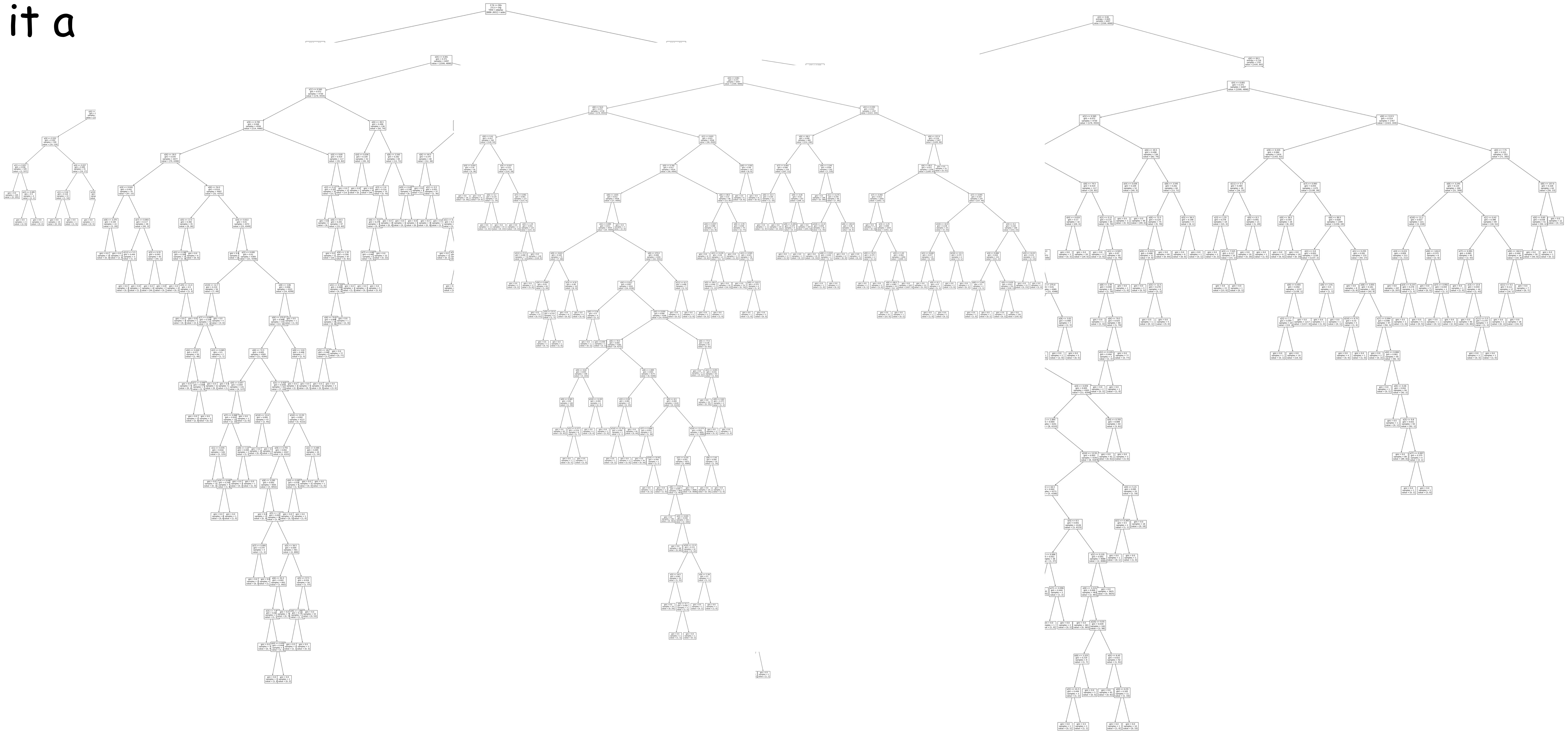
Decision Trees

1. Make splits (using features and thresholds)
2. Calculate Gini impurities
3. Select the split that results in the lowest Gini impurity sum
4. If unhappy, **just split again!**
5. Repeat 1-4 as much as needed

a hyperparameter

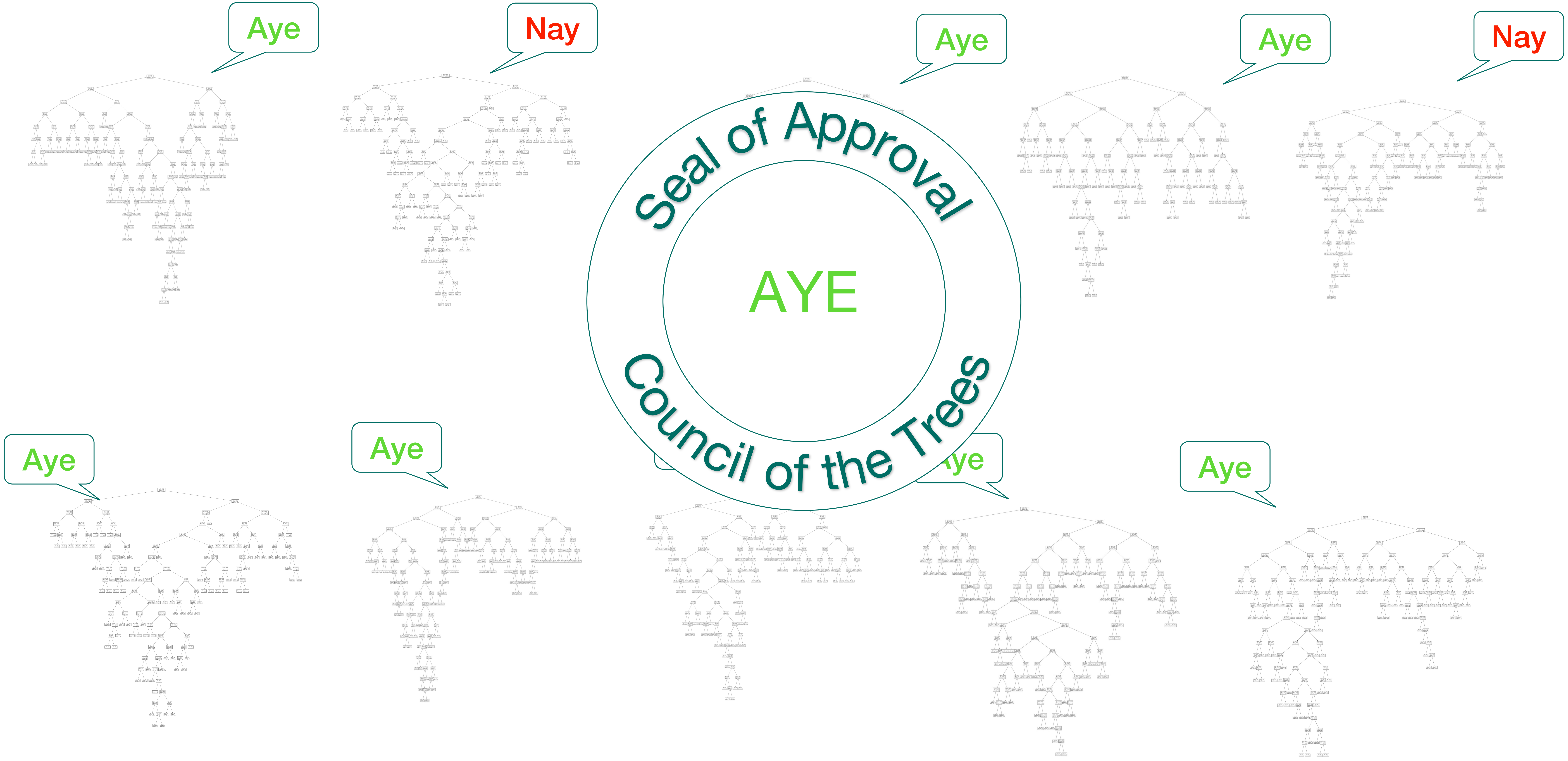
Decision Trees

diverse



What if we do it a lot?

Decision Trees
Random Forest

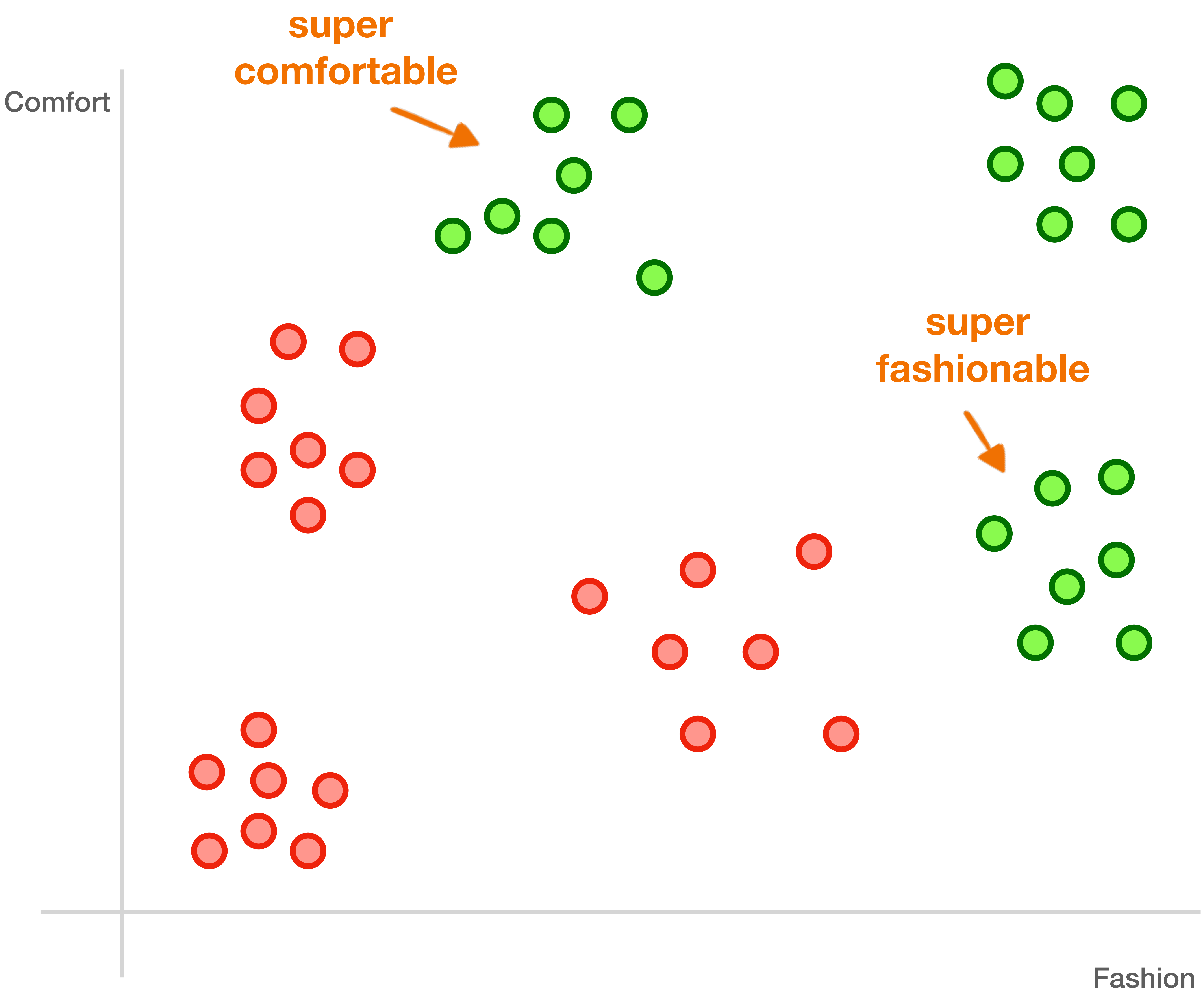


Random Forest

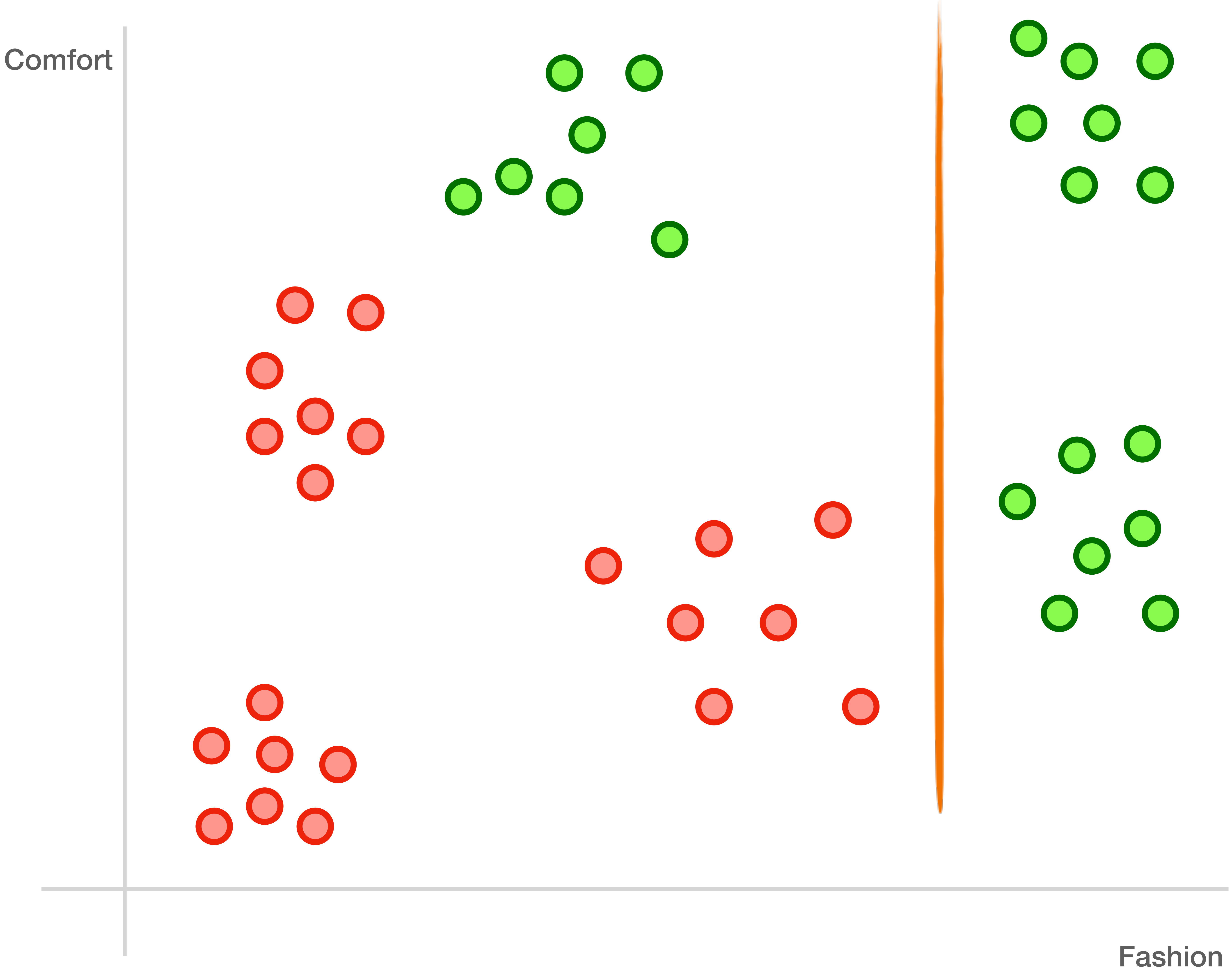
1. Make a lot of decision trees, on different portions of the data
2. For a new sample, run all of them
3. Combine their votes and take the majority

“split”

“split”

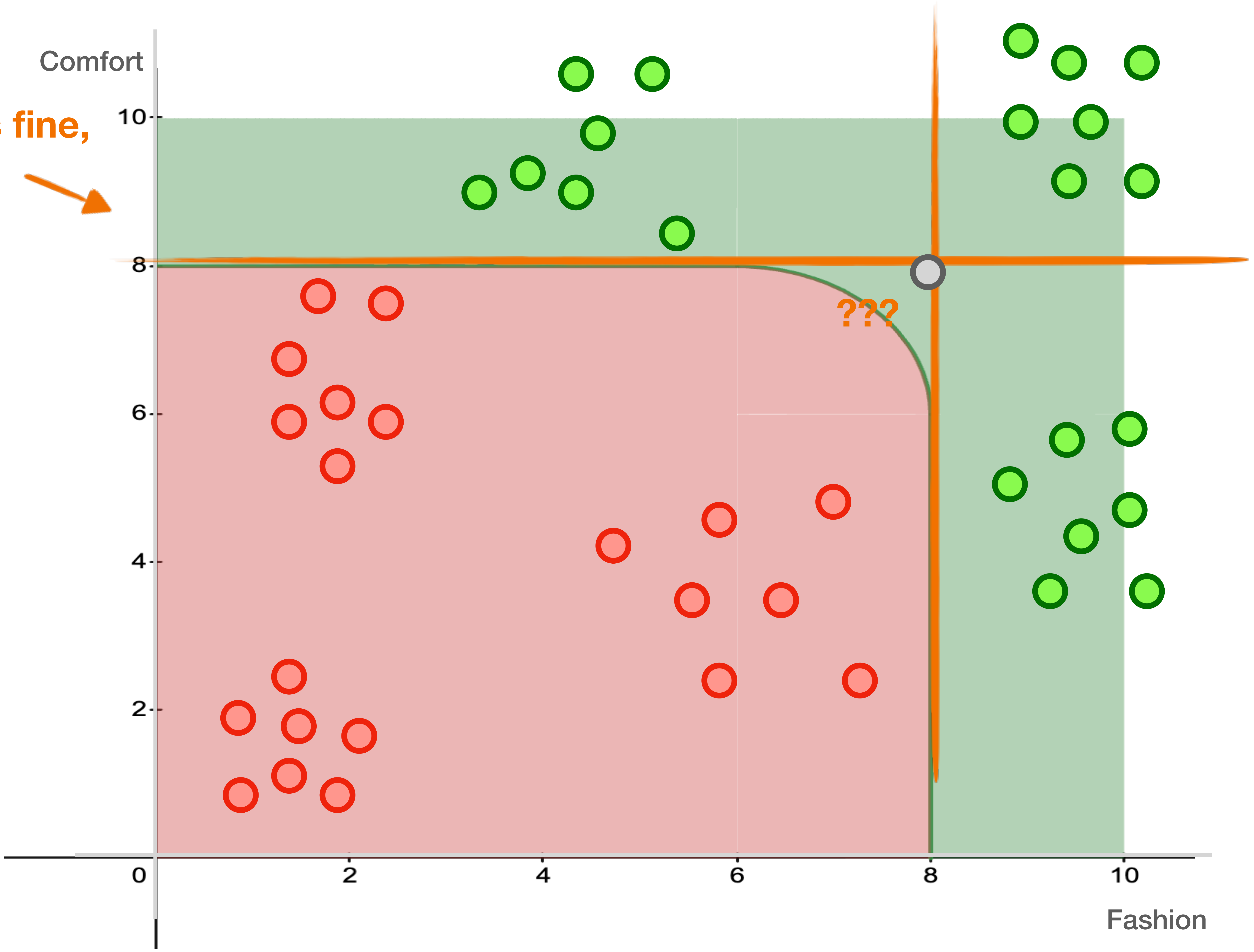


“split”

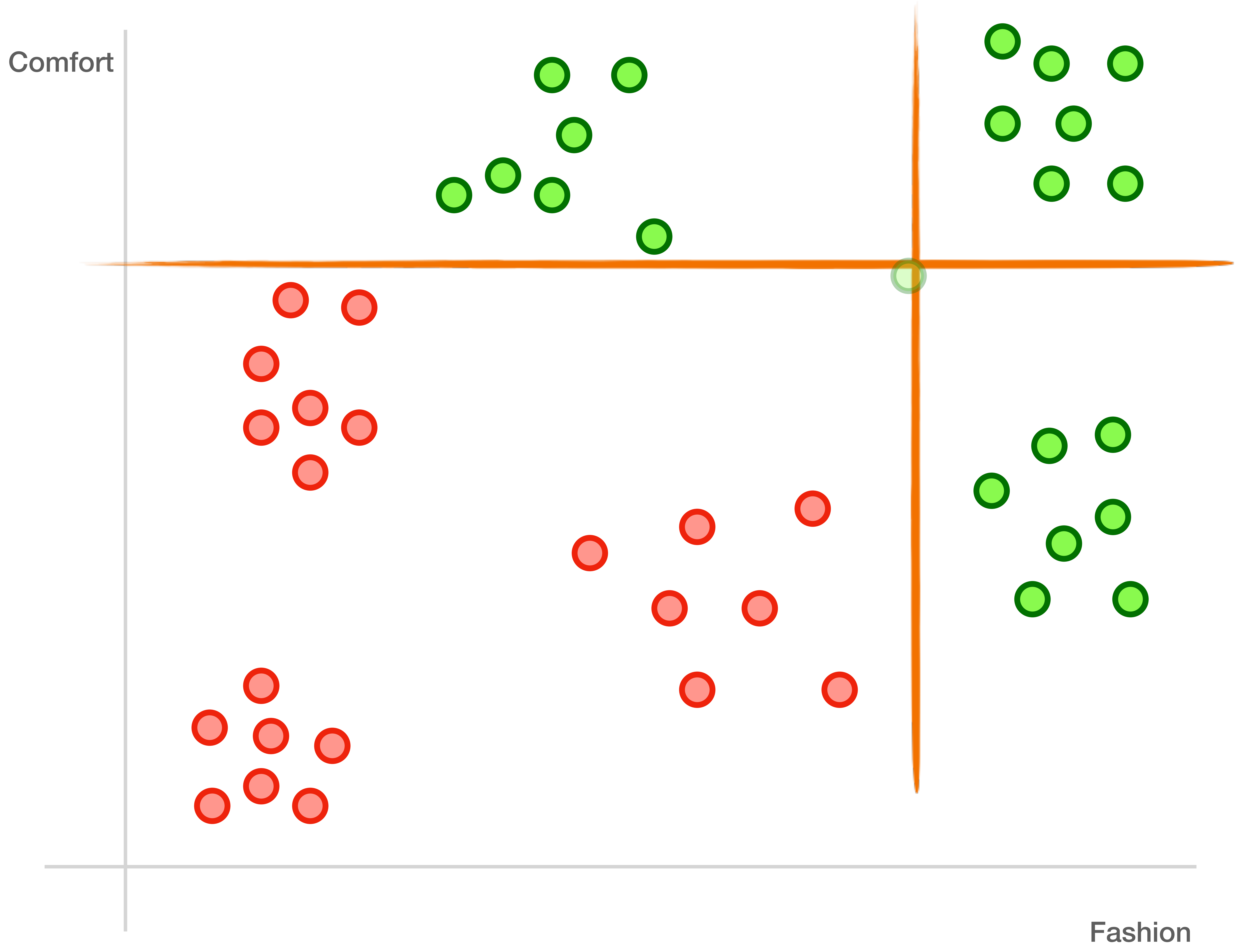


“split”

this is fine, but…



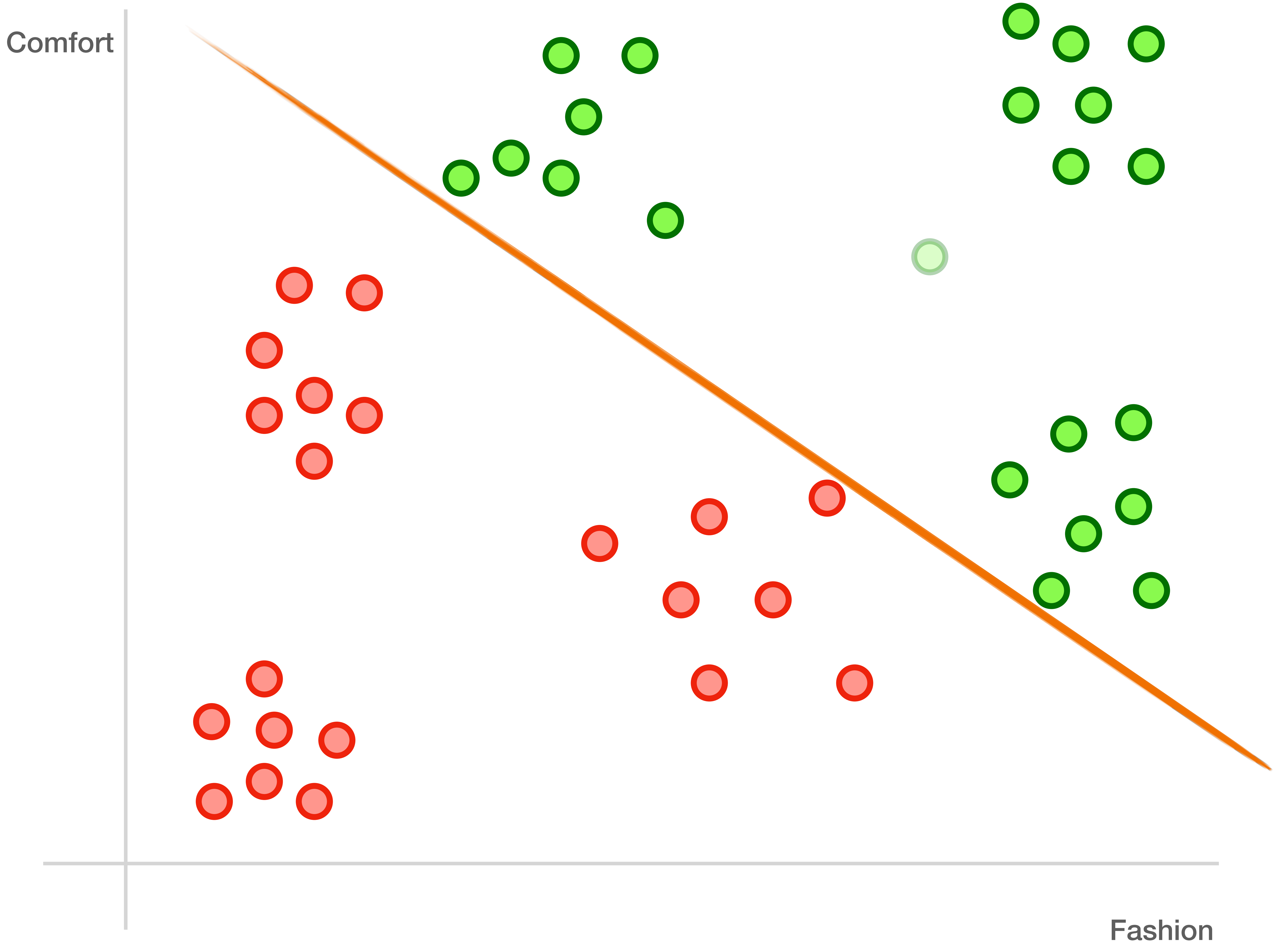
“split”



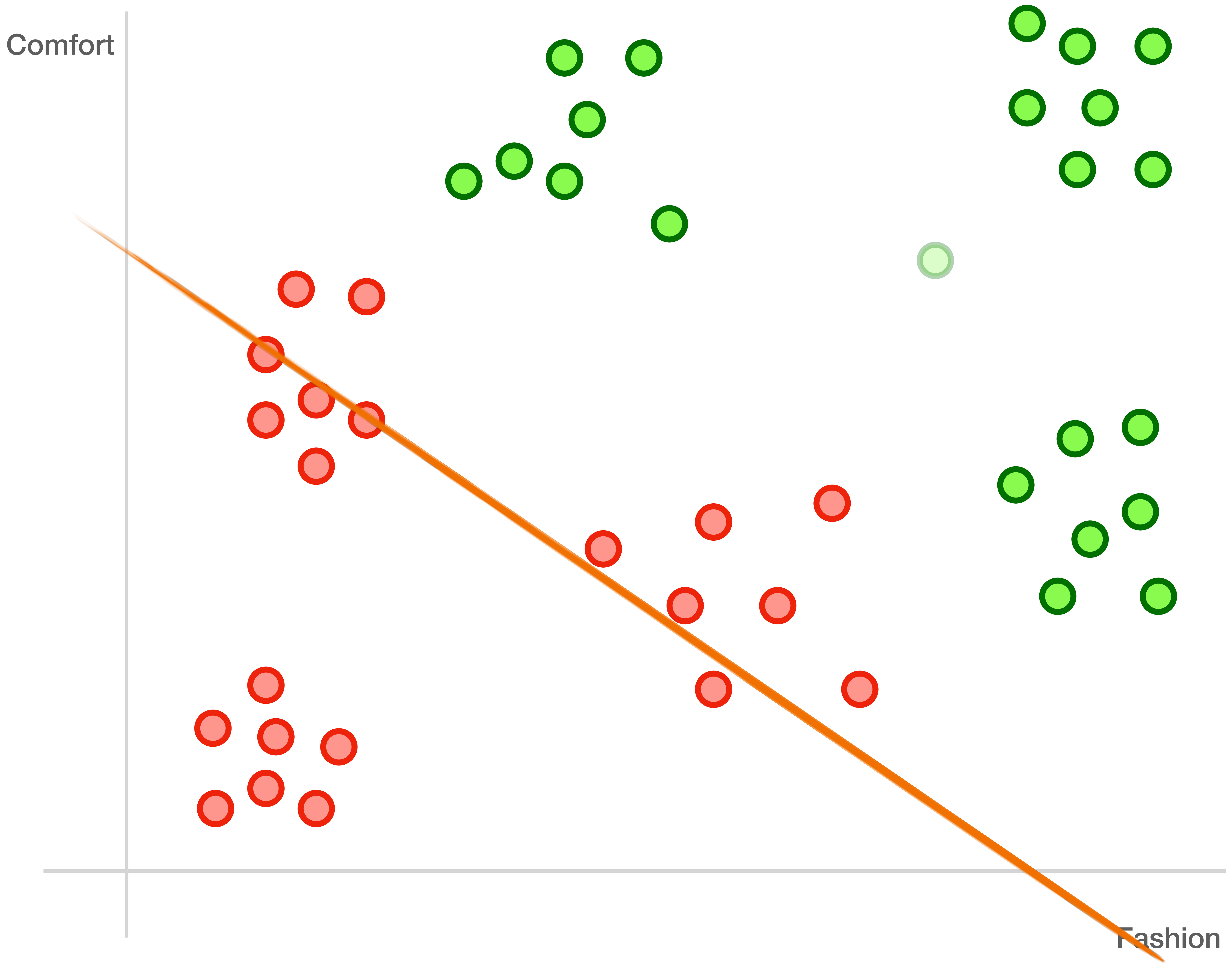
we need a more complex split

Support vector machines!

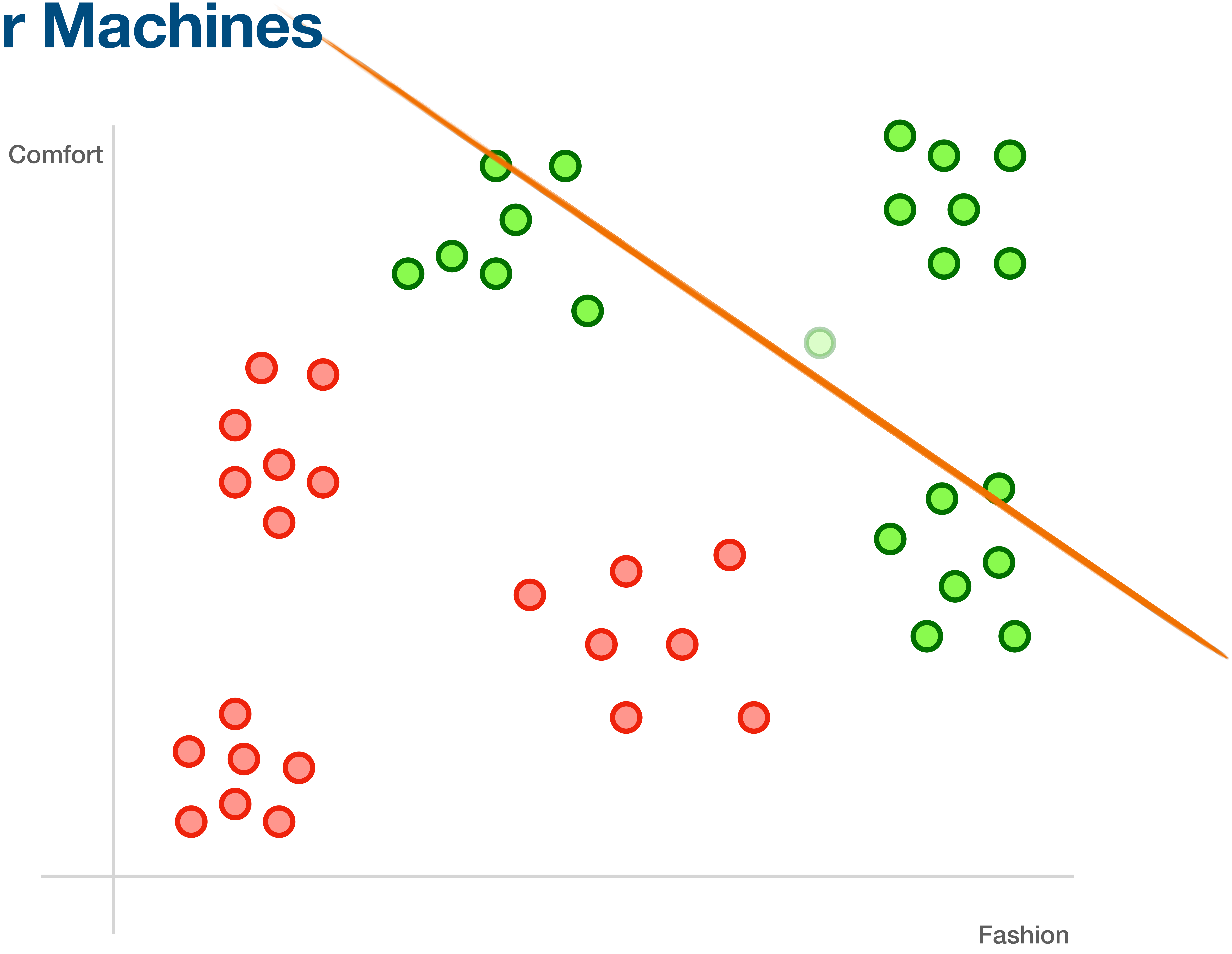
Support Vector Machines



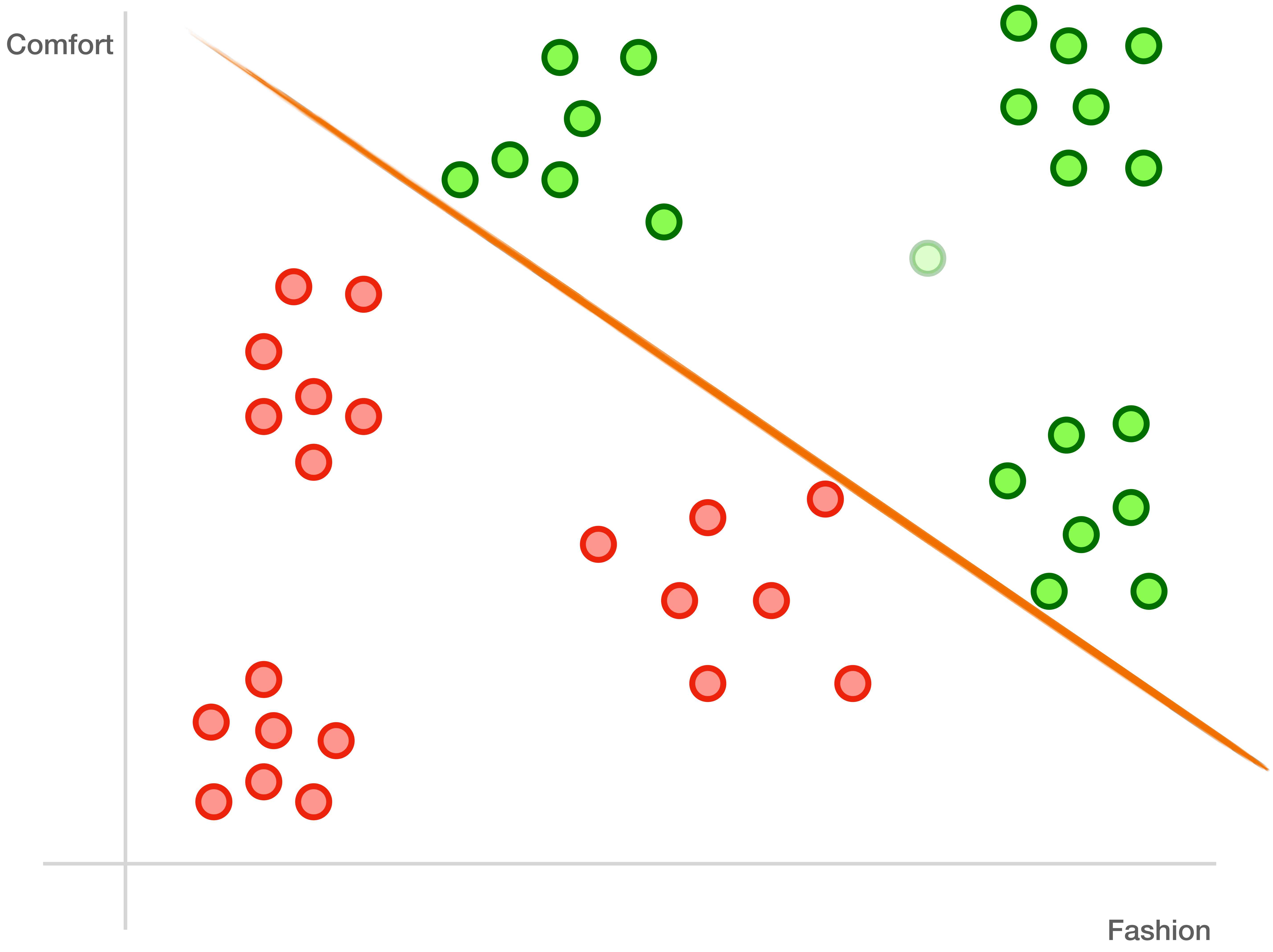
Support Vector Machines

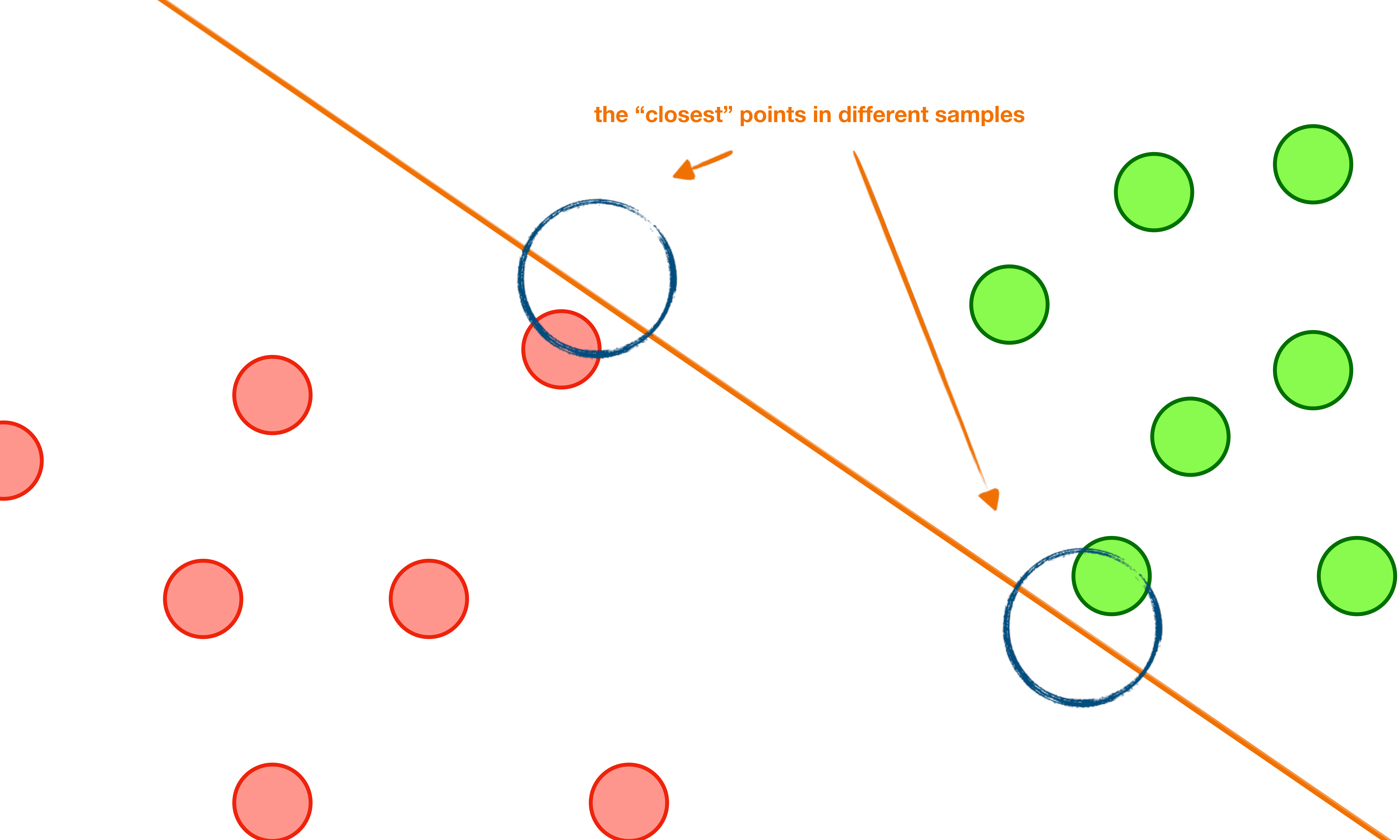


Support Vector Machines

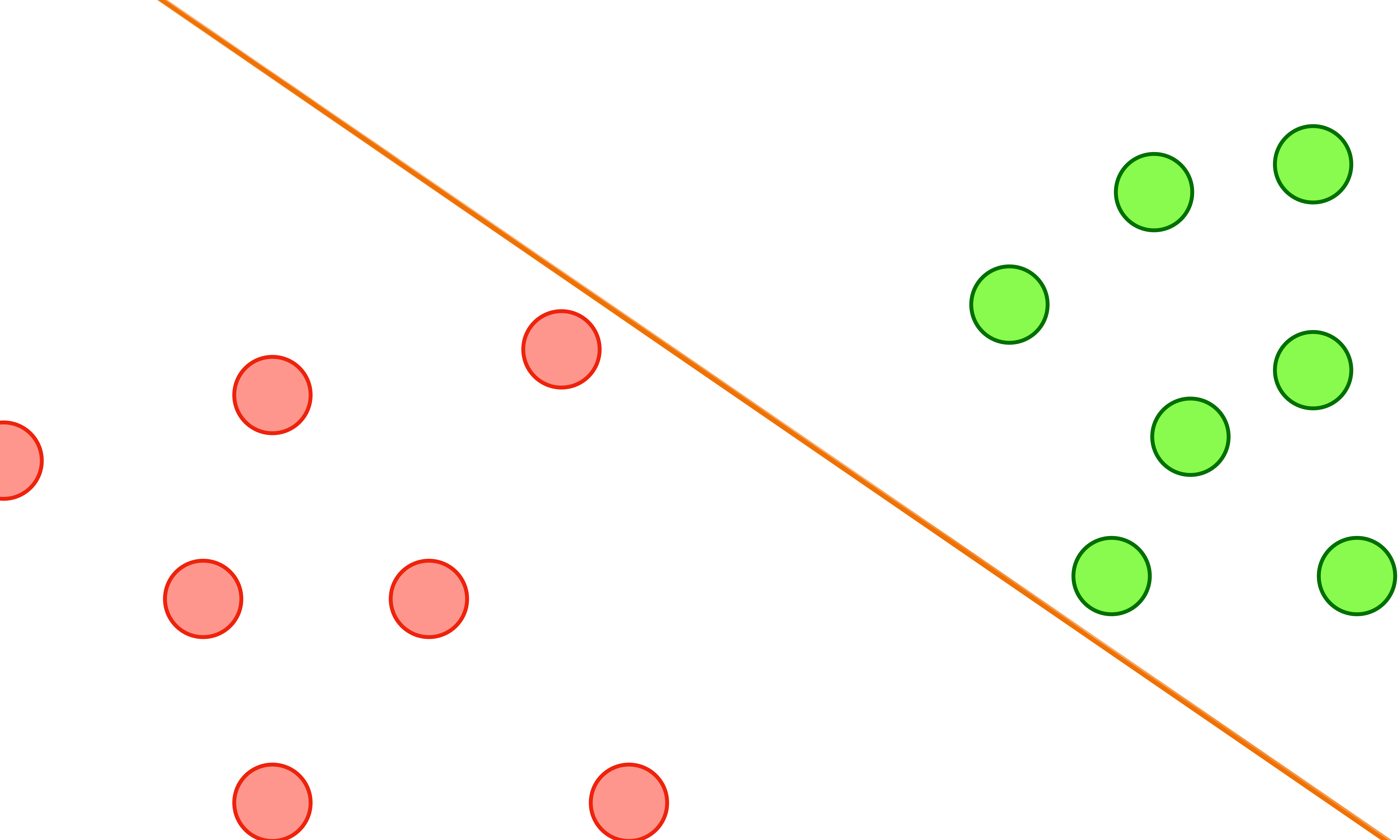


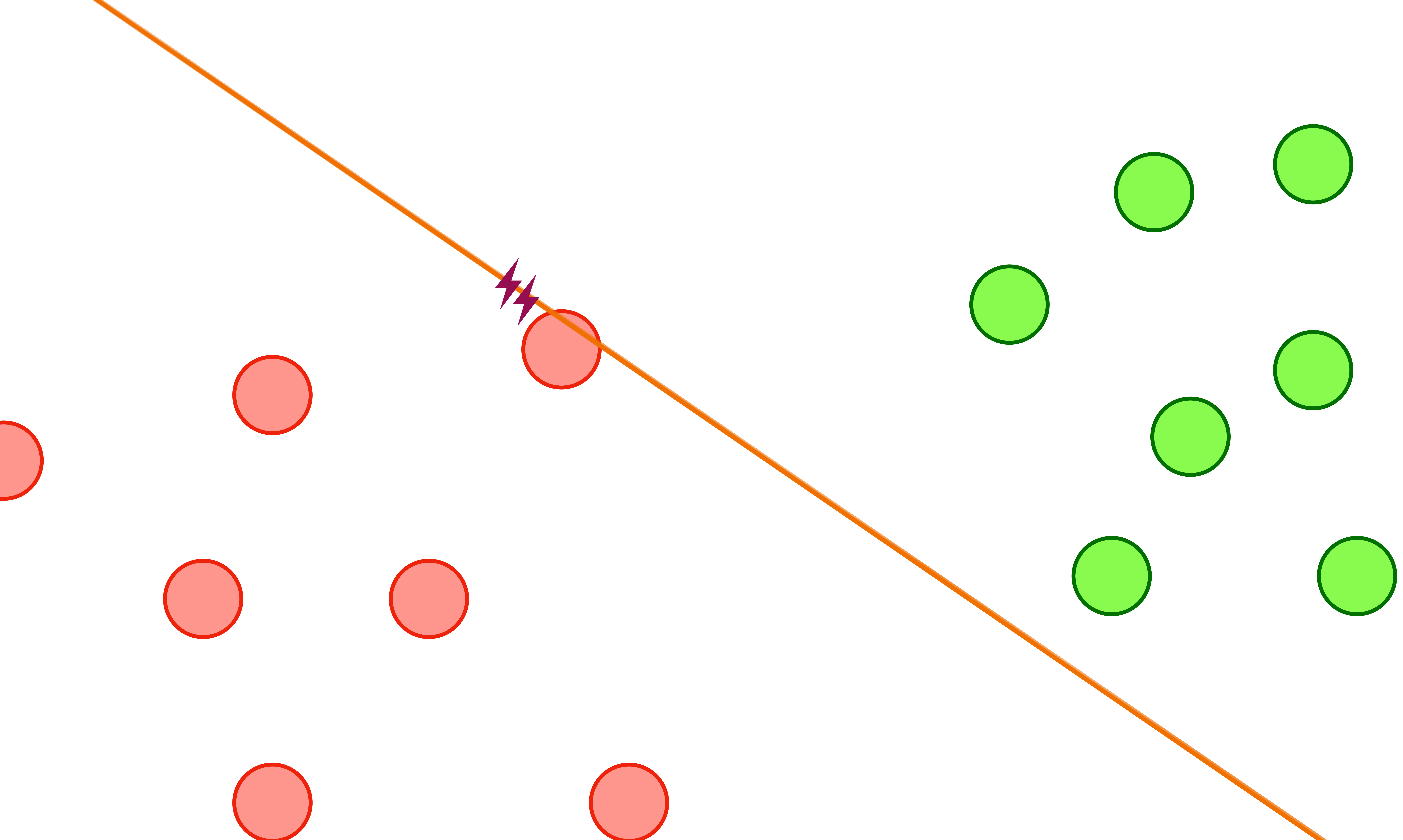
Support Vector Machines

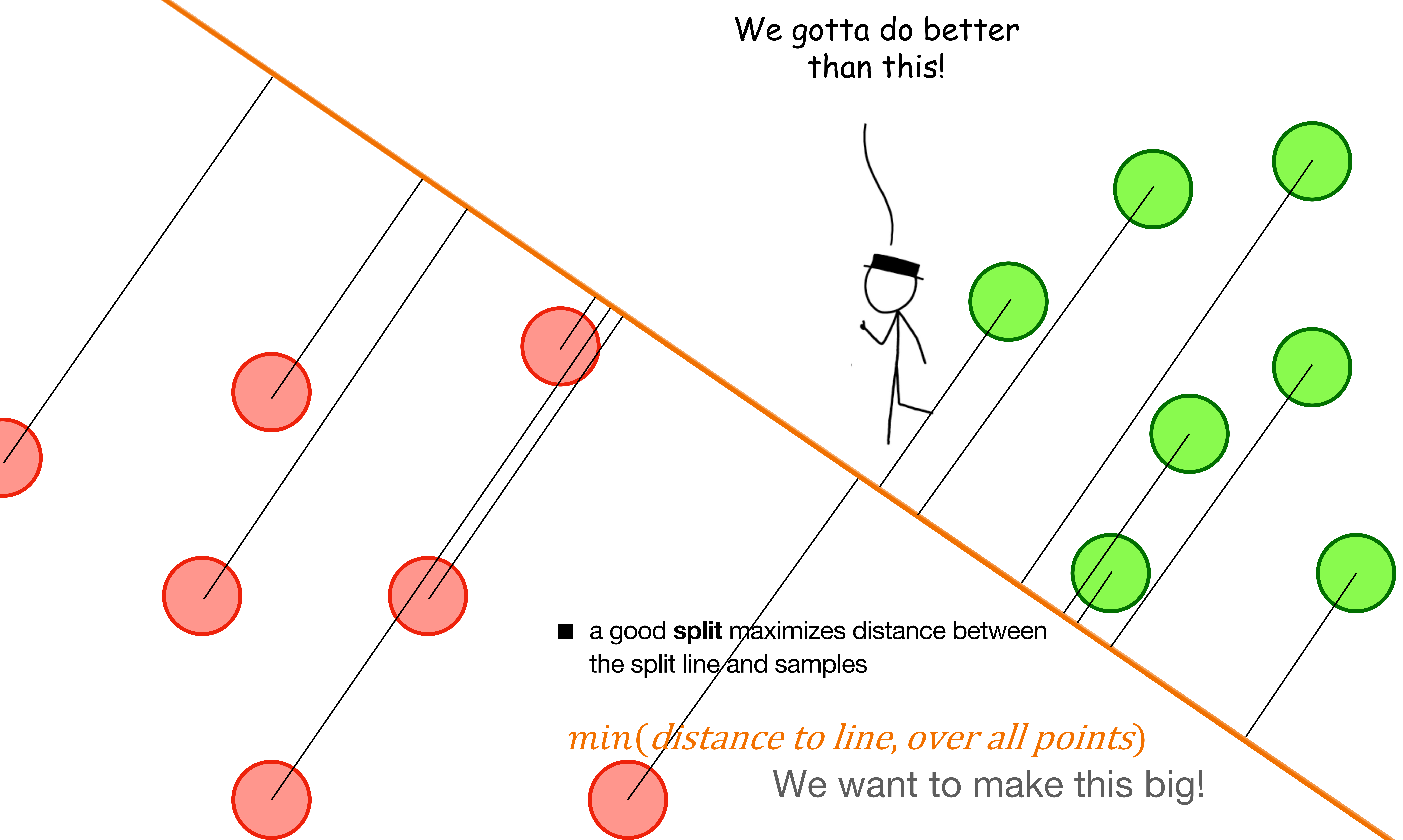




the “closest” points in different samples







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

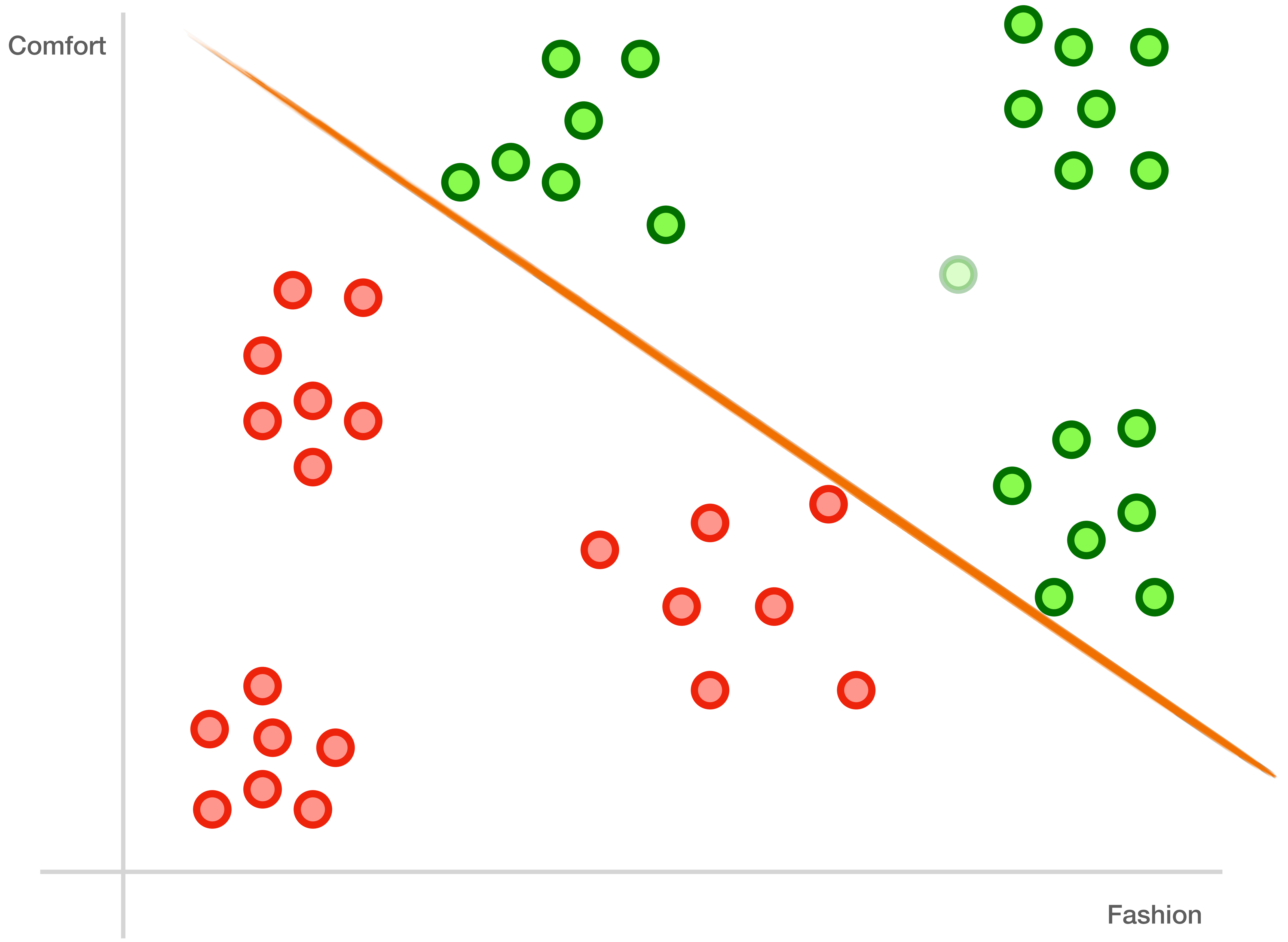
min(distance to line, over all points)

We want to make this big!

Support Vector Machines

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

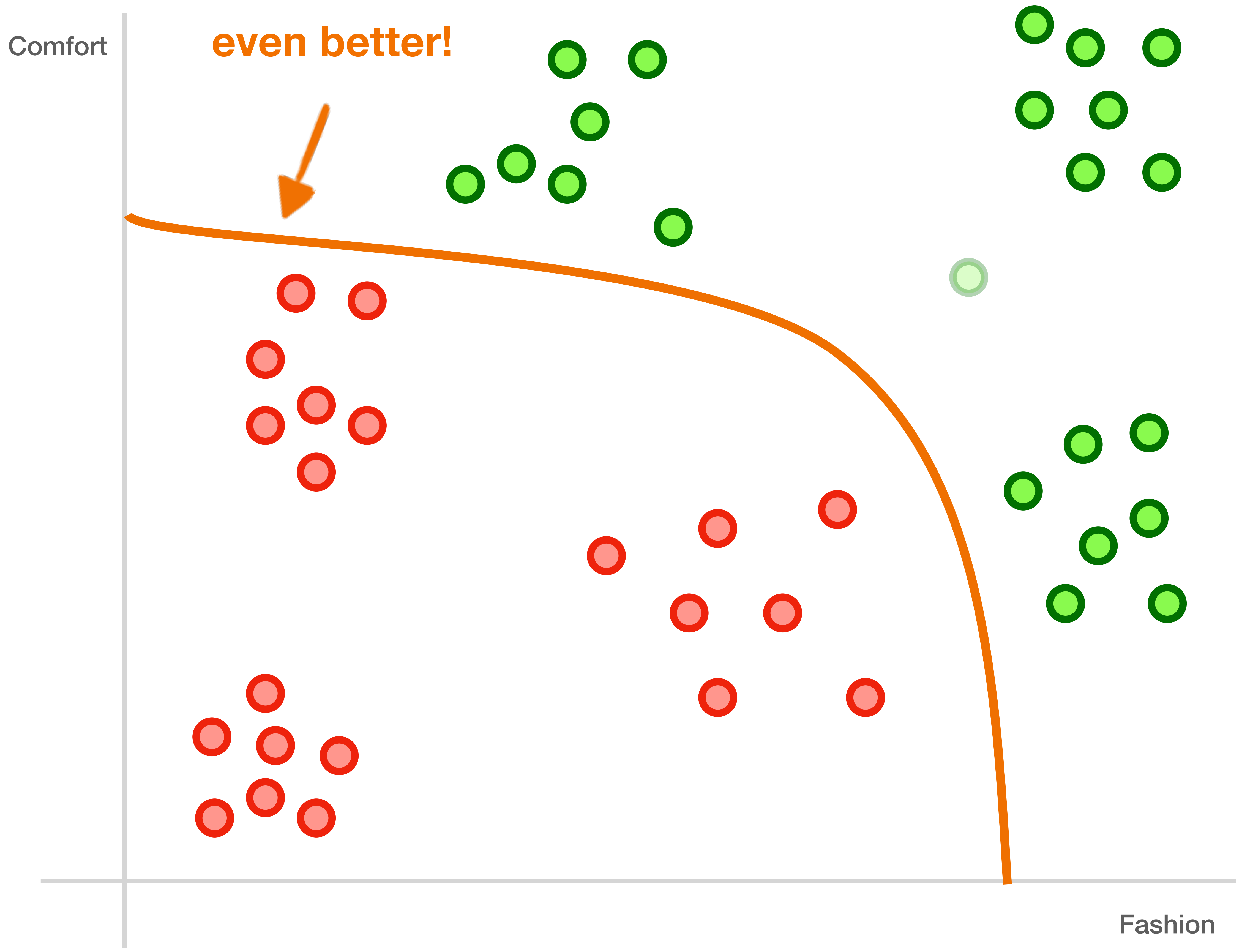
We want to make this big!



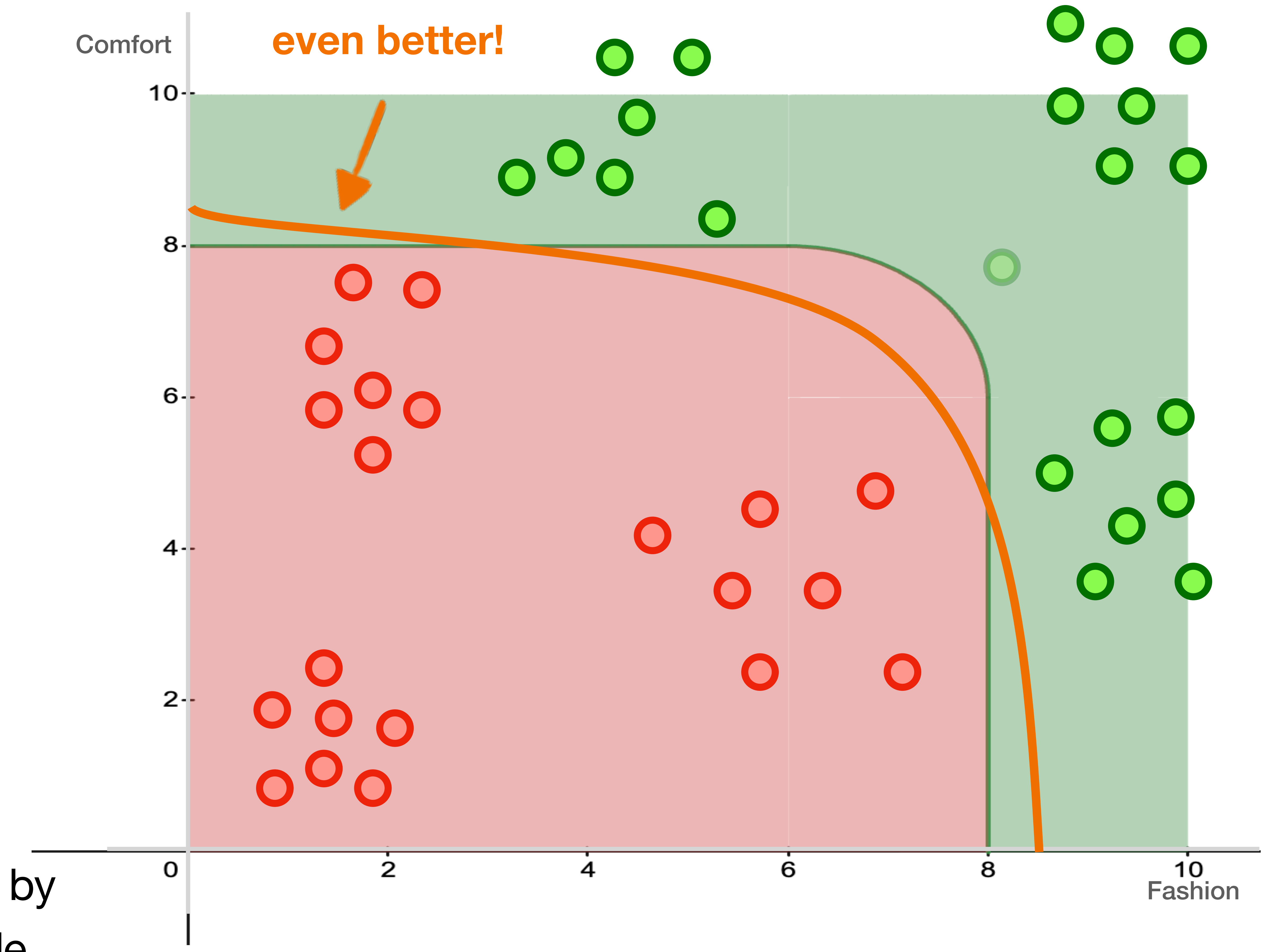
Support Vector Machines

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

We want to make this big!



Support Vector Machines



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

We want to make this big!

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

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.

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

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

we get this IF we have this

we get this IF we have this

$$A|B$$

we get this IF we have this

$$A|B$$

- Is Spam
- “Nigerian Prince”

we get this IF we have this

$$spam|nigerianprince$$

we get this IF we have this

$$P(spam|nigerianprince)$$

high? Nigerian prince → spam likely

low? Nigerian prince → not spam

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

Naïve Bayes

% of spam in dataset that relates to Nigerian prince

Bayes Equation

% of spam in dataset

$$P(spam|nigerianprince) = \frac{P(spam)P(nigerianprince|spam)}{P(nigerianprince)}$$

% of Nigerian prince in dataset

Naïve Bayes

Classifier

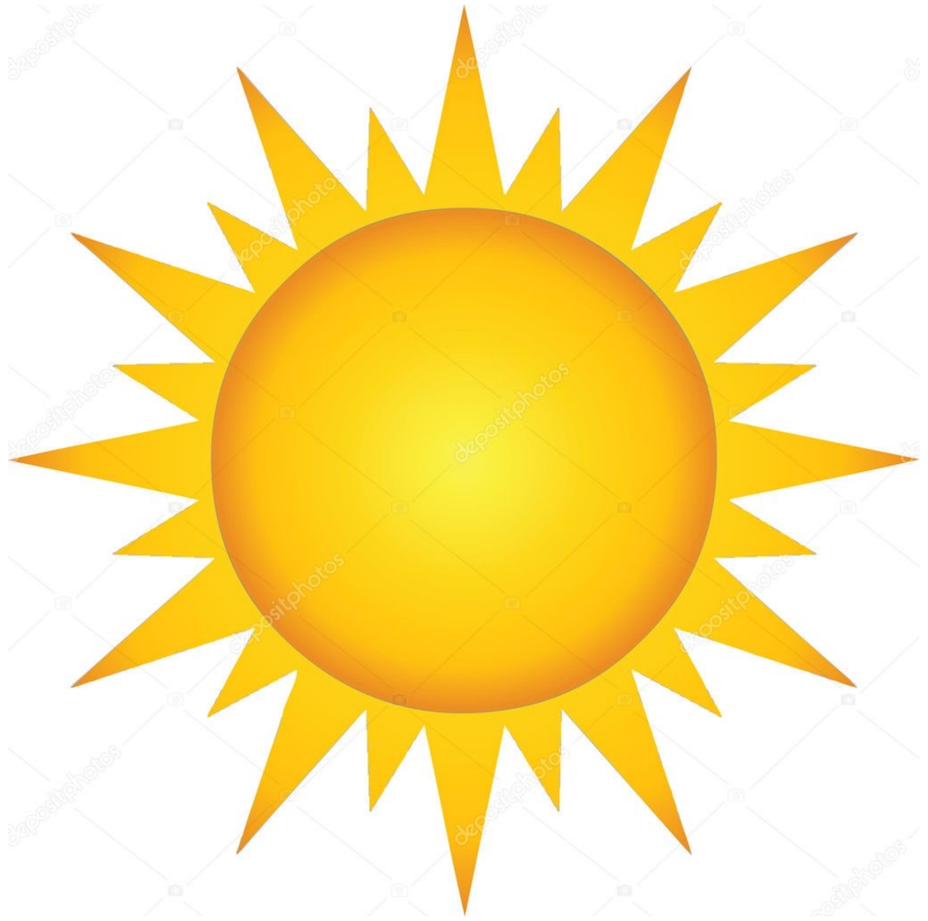
$$P(spam|nigerianprince, offer) = \frac{P(spam)P(nigerianprince|spam)P(offer|spam)}{P(nigerianprince)P(offer)}$$

- **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"

Naïve Bayes
Independence



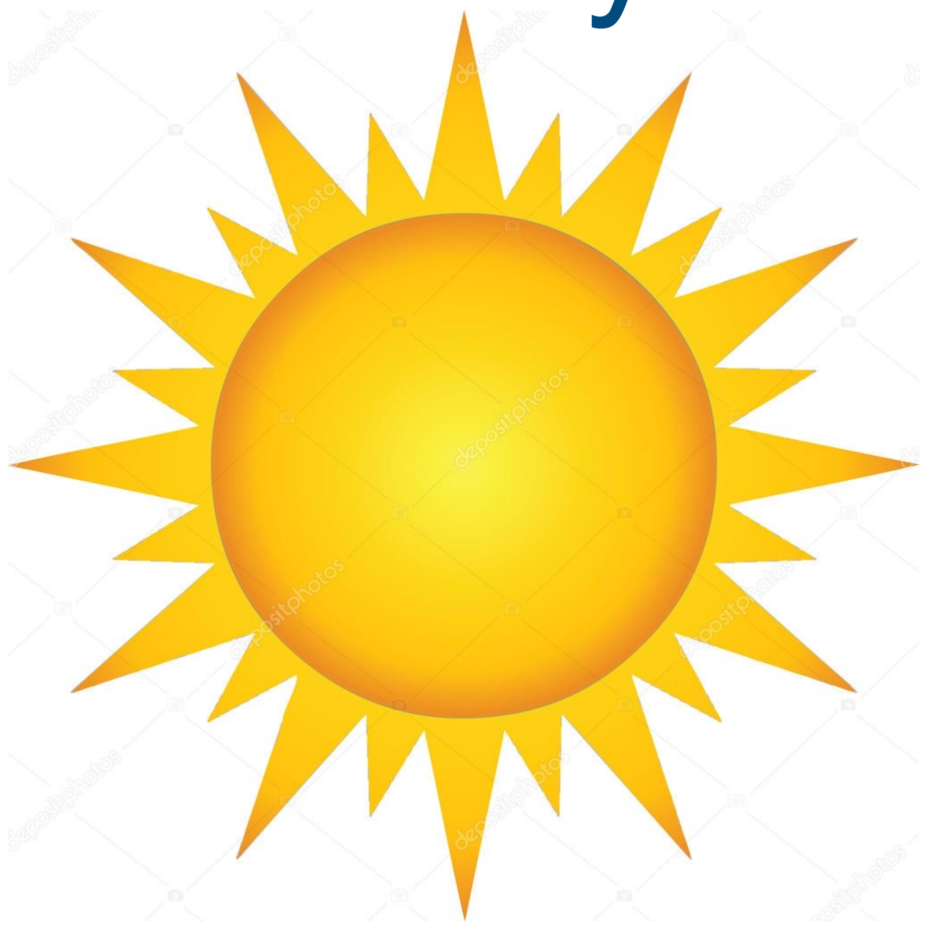
50%

50%

Naïve Bayes

Independence

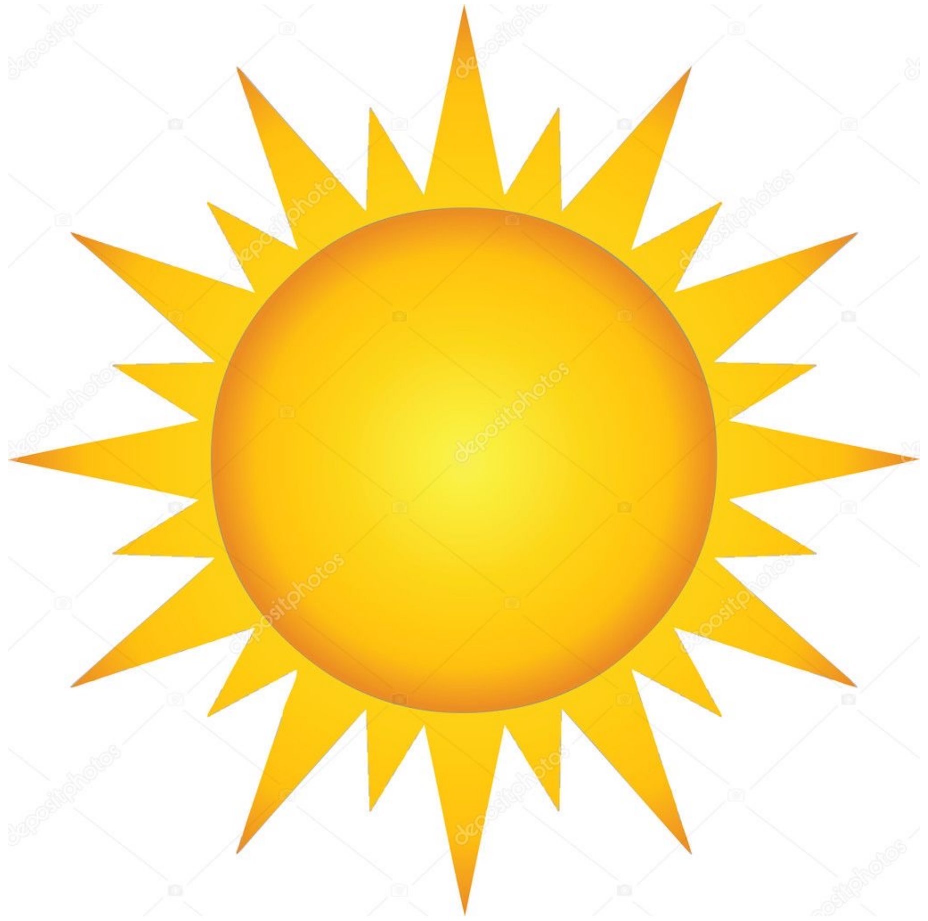
January 1st



50%

50%

January 2nd



50%

50%

Naïve Bayes

Independence

P(Rain | January 1st)

=50%

P(Rain | January 2nd)

=50%

Naïve Bayes
Independence

$P(\text{Rain} \mid \text{January } 1^{st} \text{ AND } \text{Rain} \mid \text{January } 2^{nd}) \quad = 45\%$

Is NOT

$P(\text{Rain} \mid \text{January } 1^{st}) * P(\text{Rain} \mid \text{January } 2^{nd})$
$= 25\%$

Naïve Bayes Classifier

$$P(spam|nigerianprince, offer) = \frac{P(spam)P(nigerianprince|spam)P(offer|spam)}{P(nigerianprince)P(offer)}$$

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

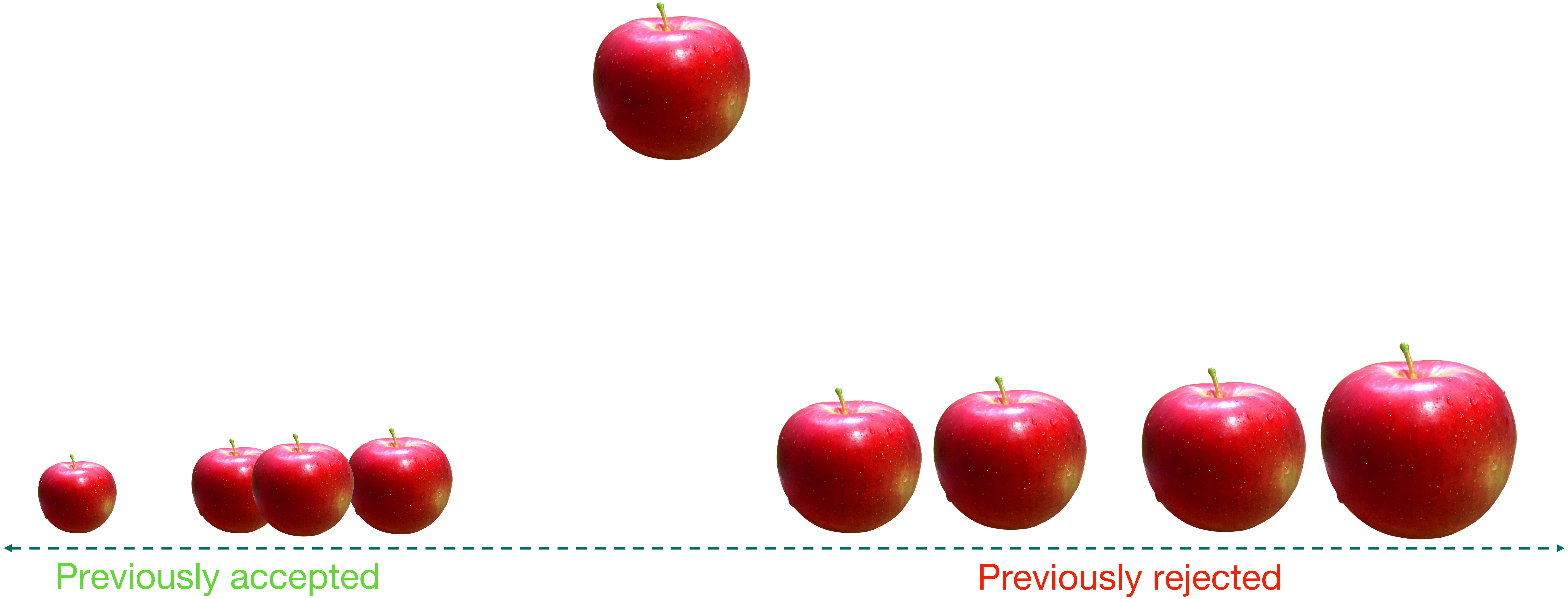
multiplication for AND assumes independence!

"naïve"

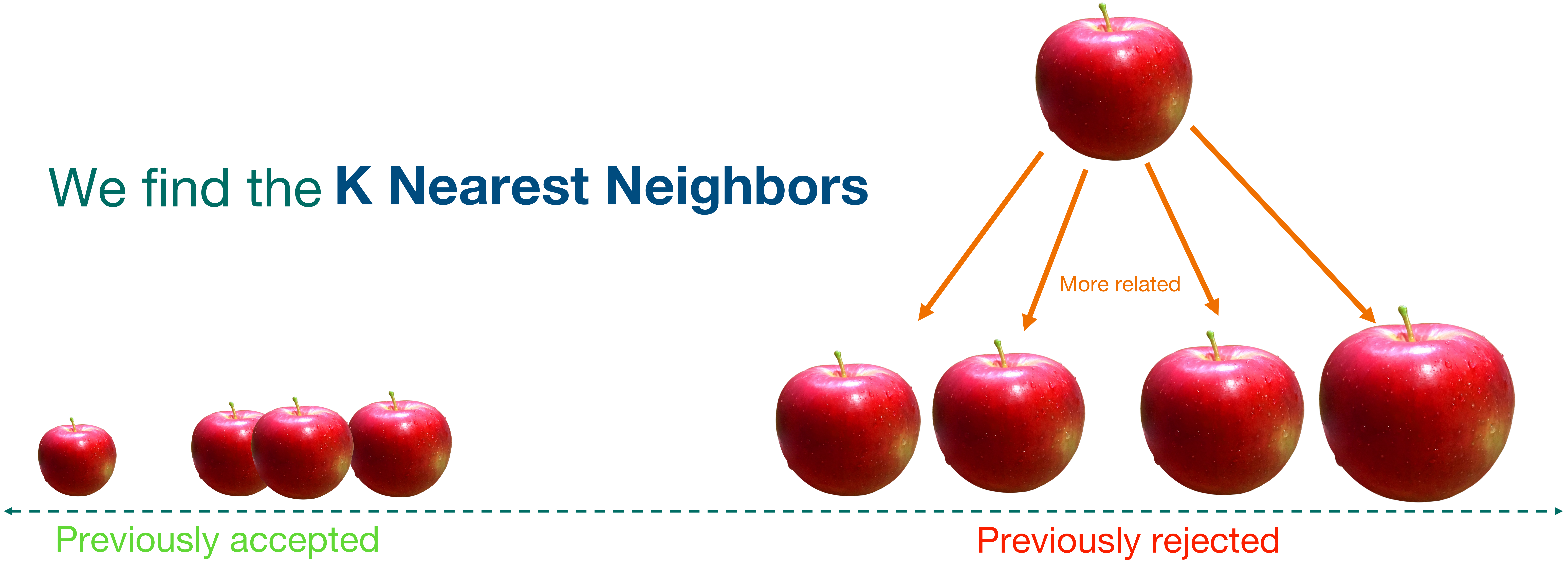
Buy? Don’t buy?

Did Buy

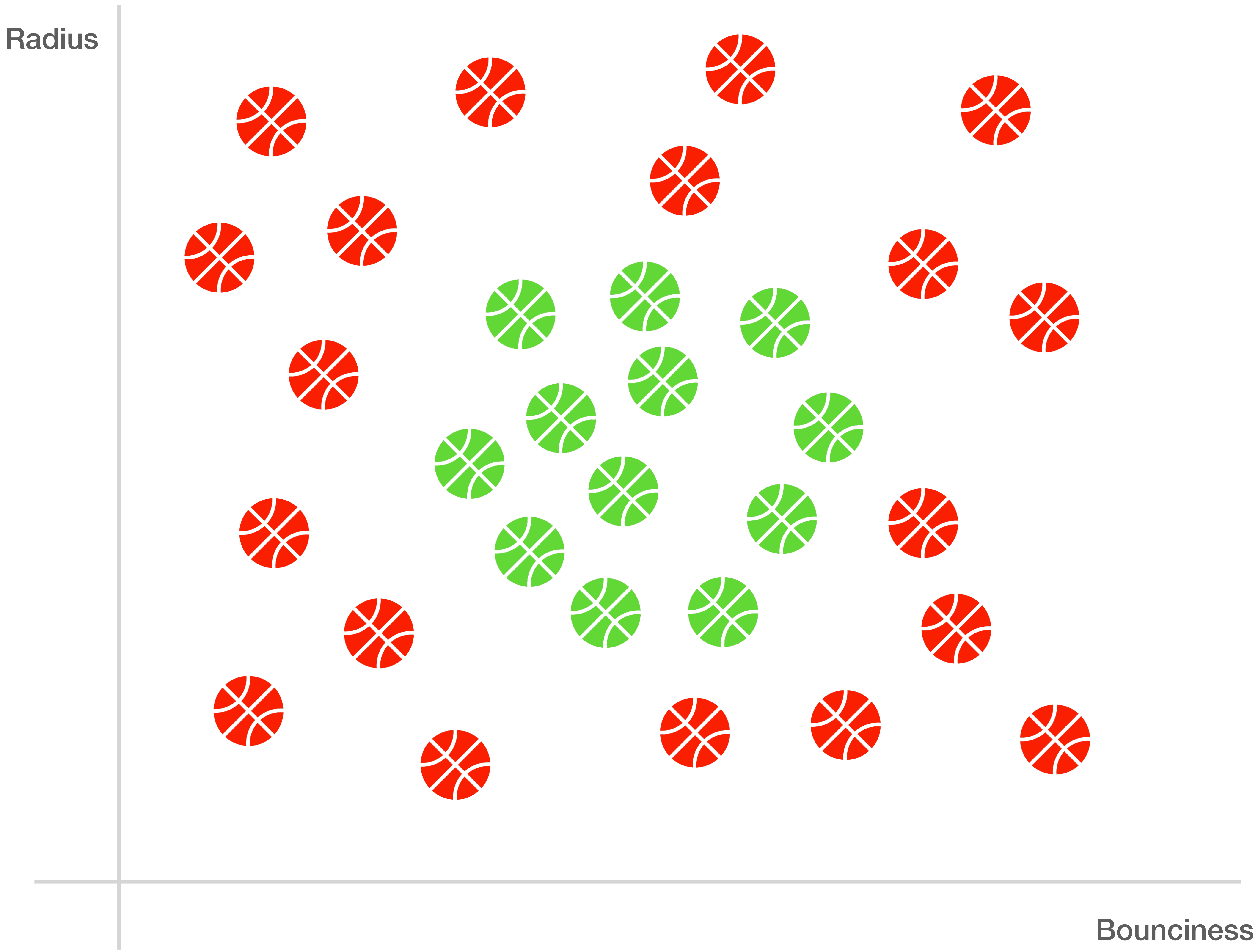
Rejected


Previously accepted
Previously rejected

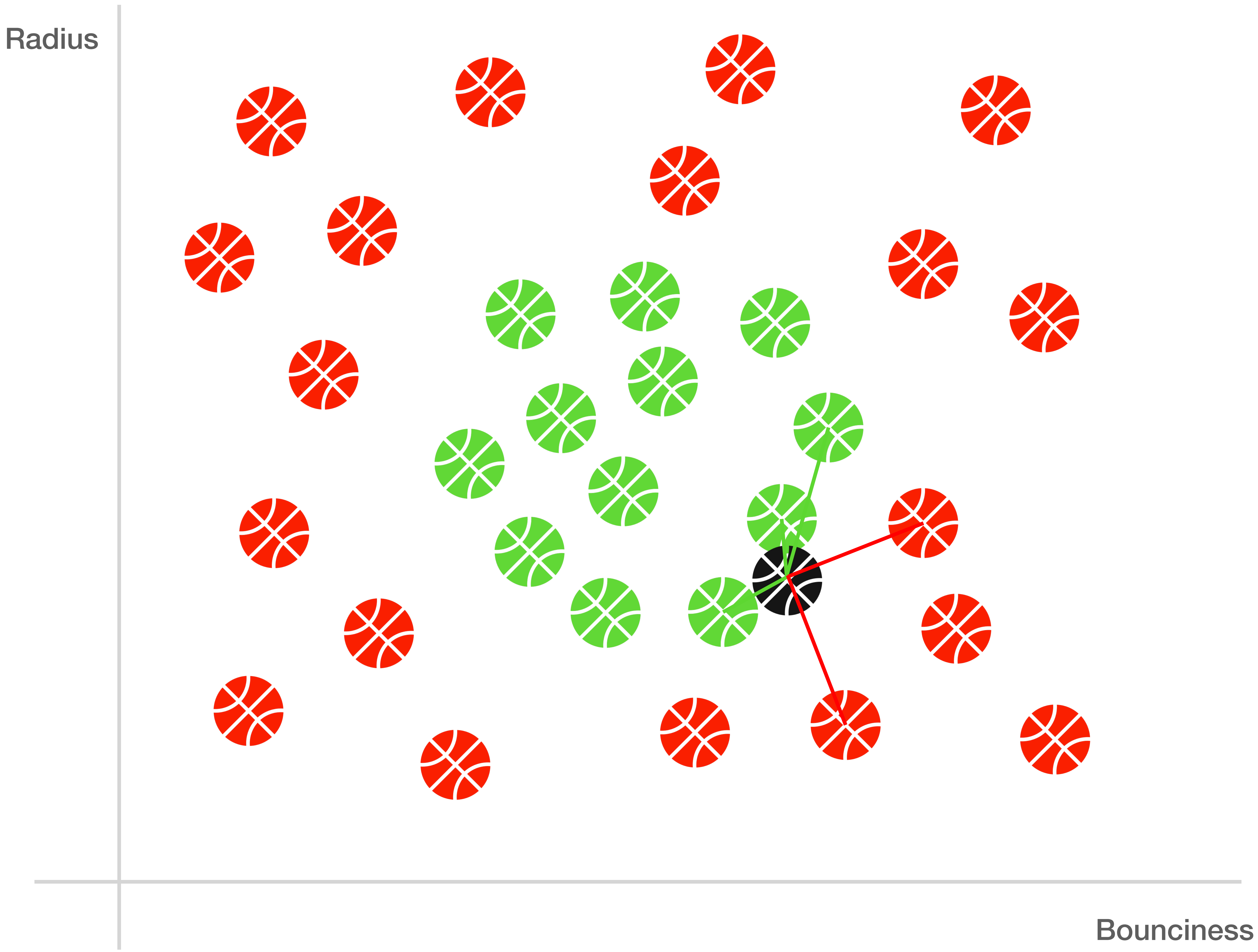
We find the **K Nearest Neighbors**



K Nearest Neighbors



K Nearest Neighbors

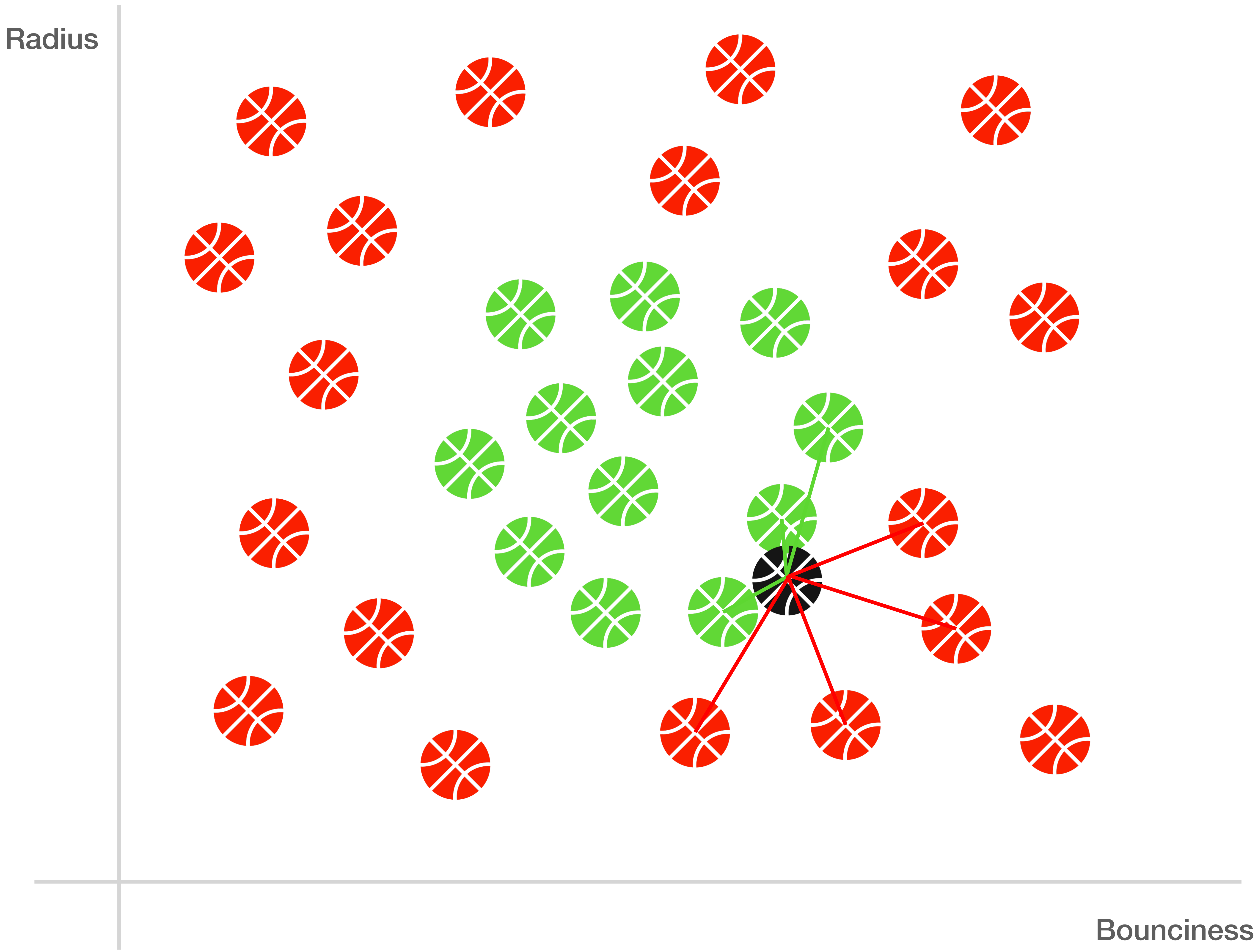


another **hyperparameter**

K=5

Buy

K Nearest Neighbors



another **hyperparameter**

K=7

Don’t Buy

Five classifiers! That’s a lot.

Let’s get to the lab!