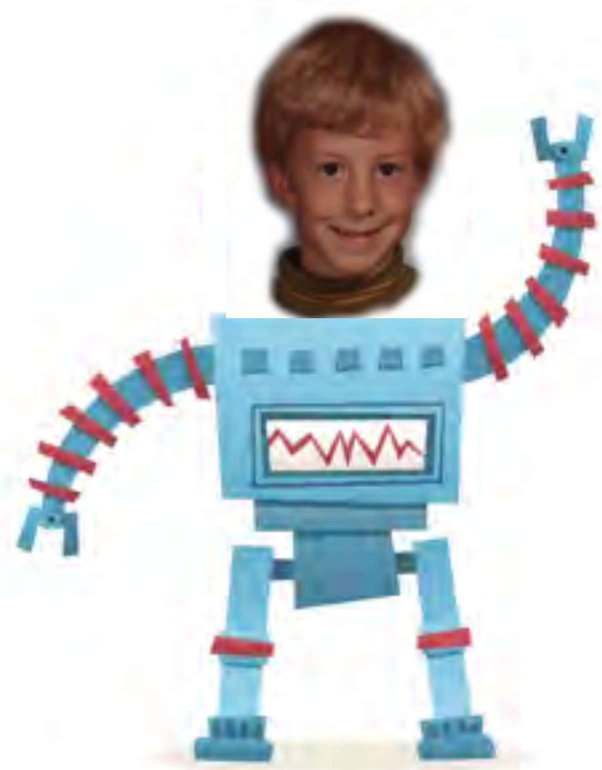
Machine Learning for Data Science

Robert Nowak
University of Wisconsin

Robot Rob

Robot Rob

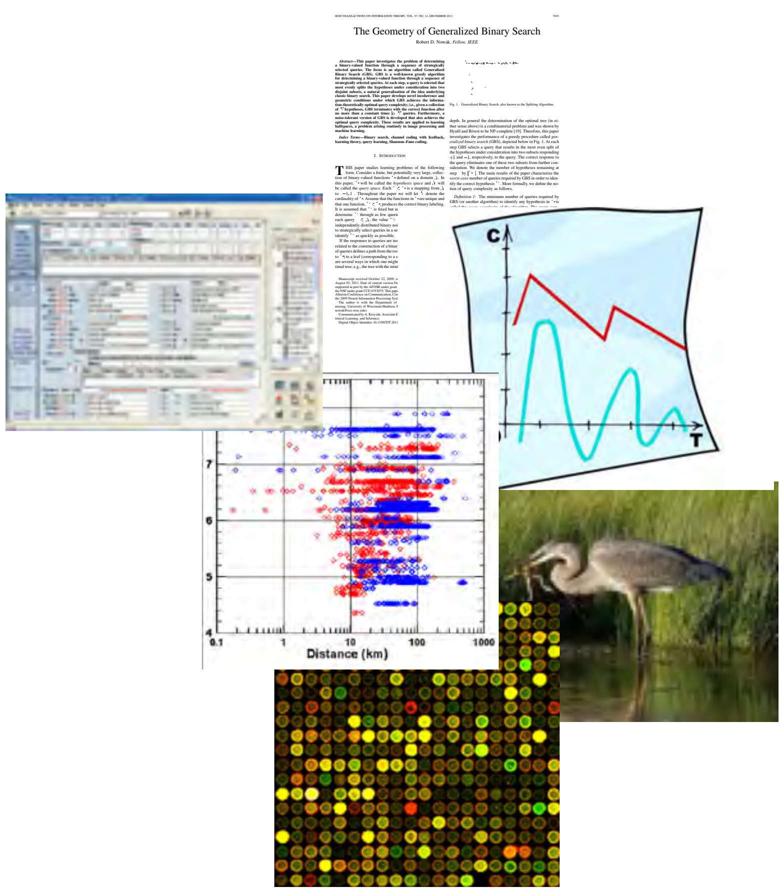


What is Machine Learning?

Machine learning is an area of Computer Science focused on designing computer programs that enable machines to learn by example, much in the way young children are taught to understand the world around them.

What is Machine Learning?

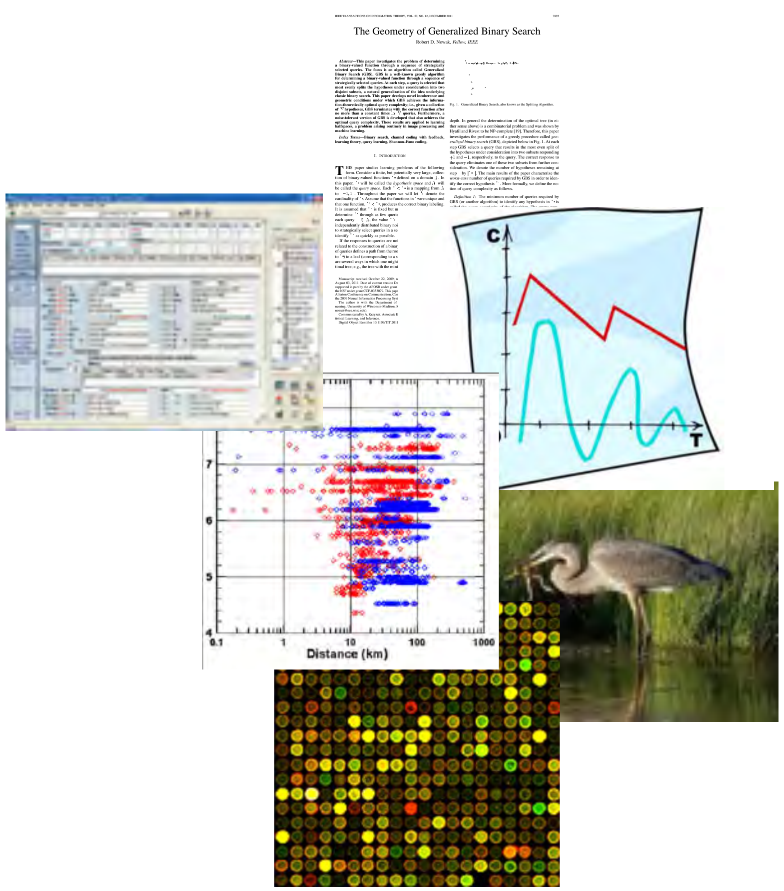
Machine learning is an area of Computer Science focused on designing computer programs that enable machines to learn by example, much in the way young children are taught to understand the world around them.



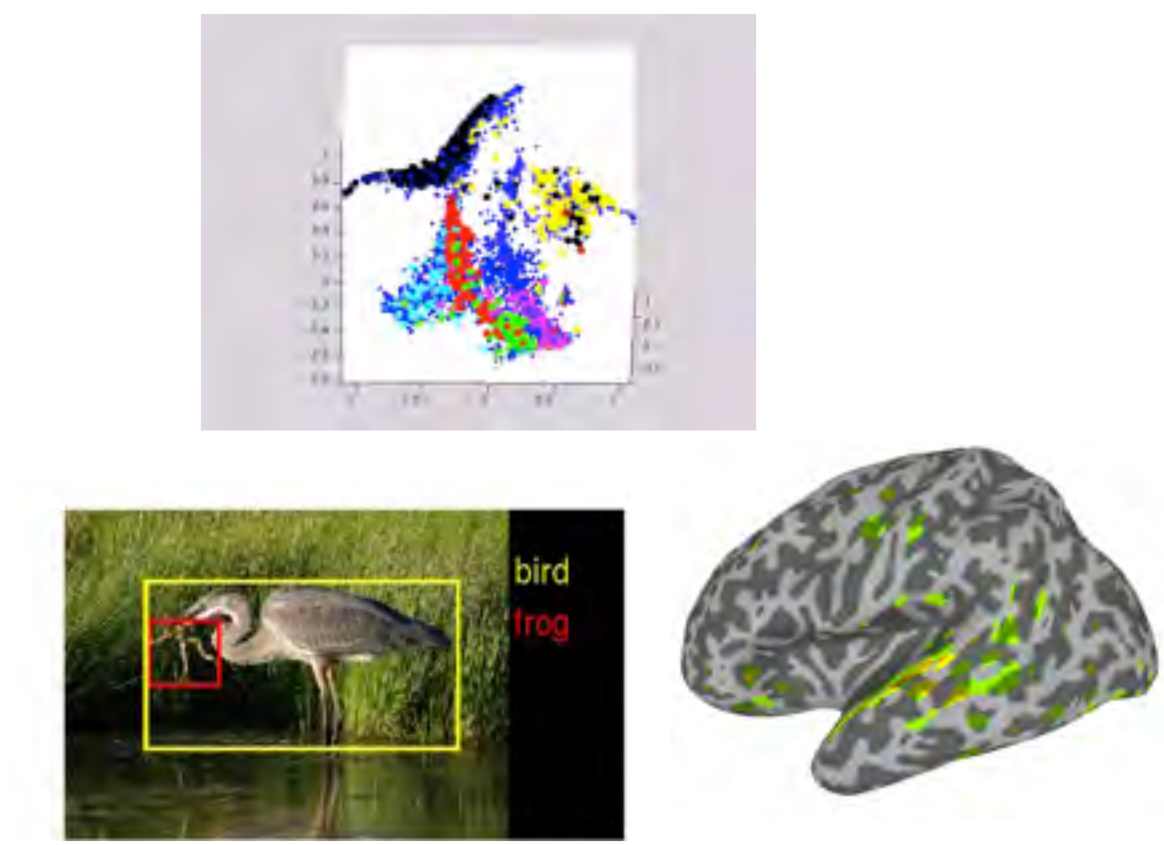
Raw Data

What is Machine Learning?

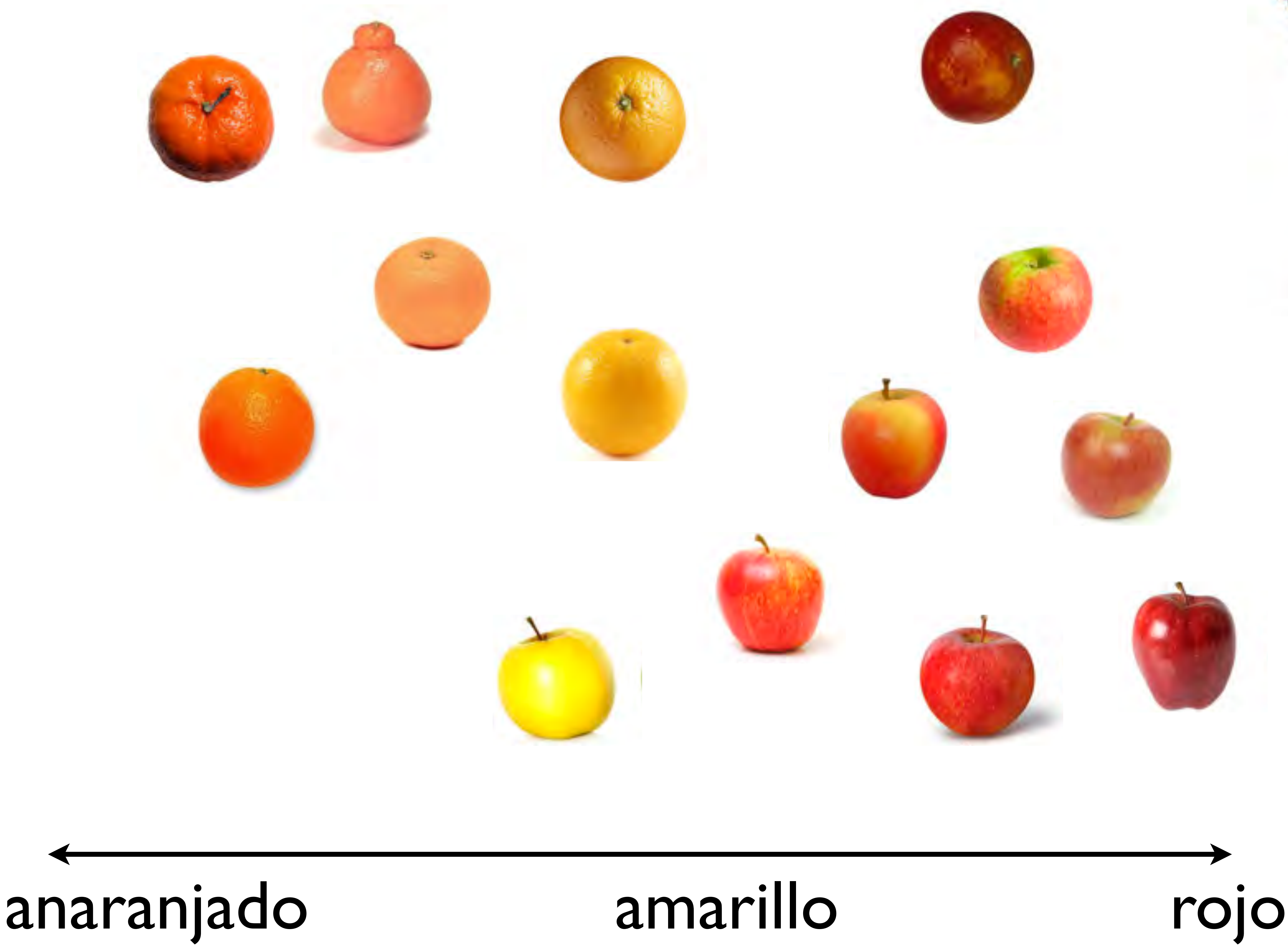
Machine learning is an area of Computer Science focused on designing computer programs that enable machines to learn by example, much in the way young children are taught to understand the world around them.

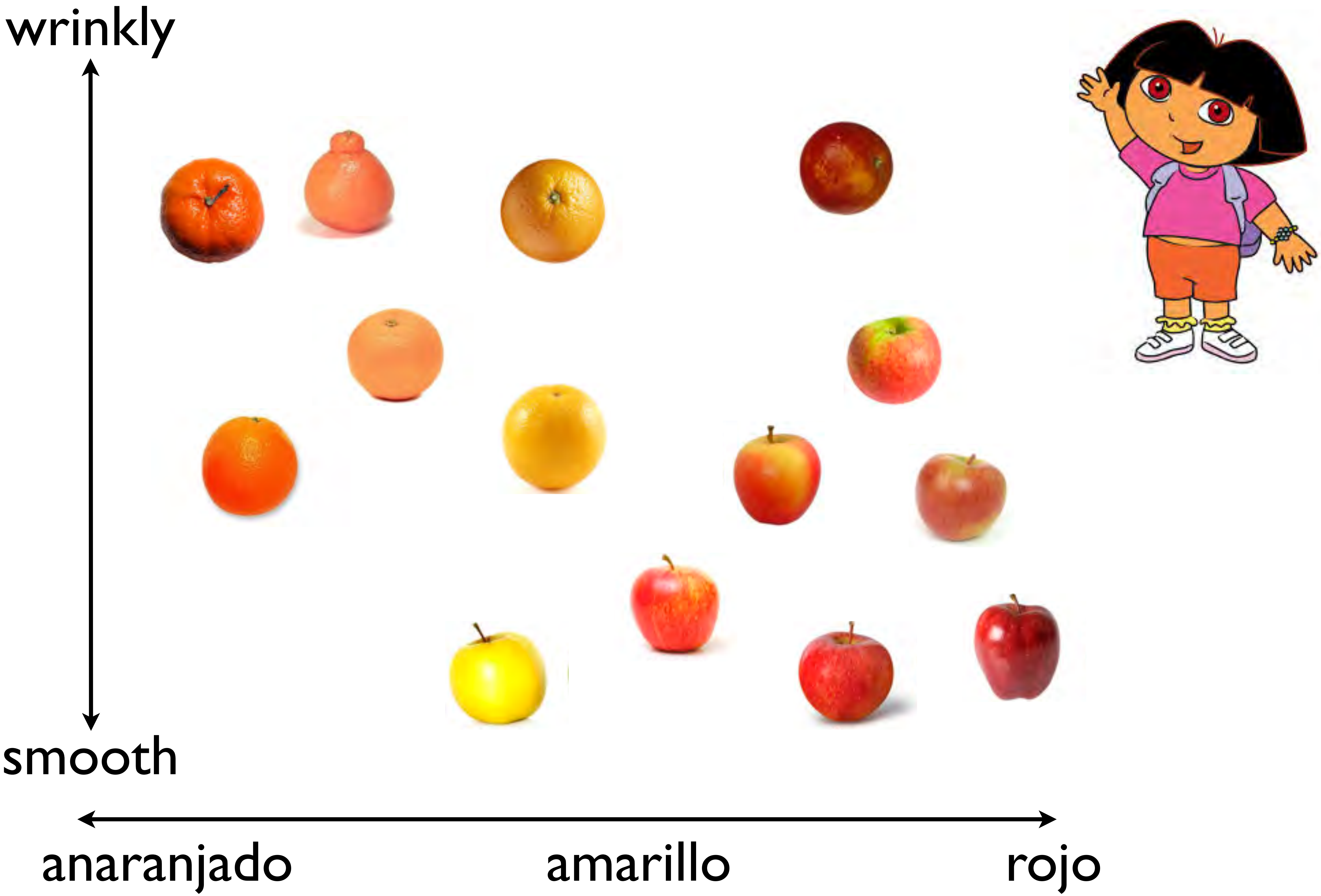


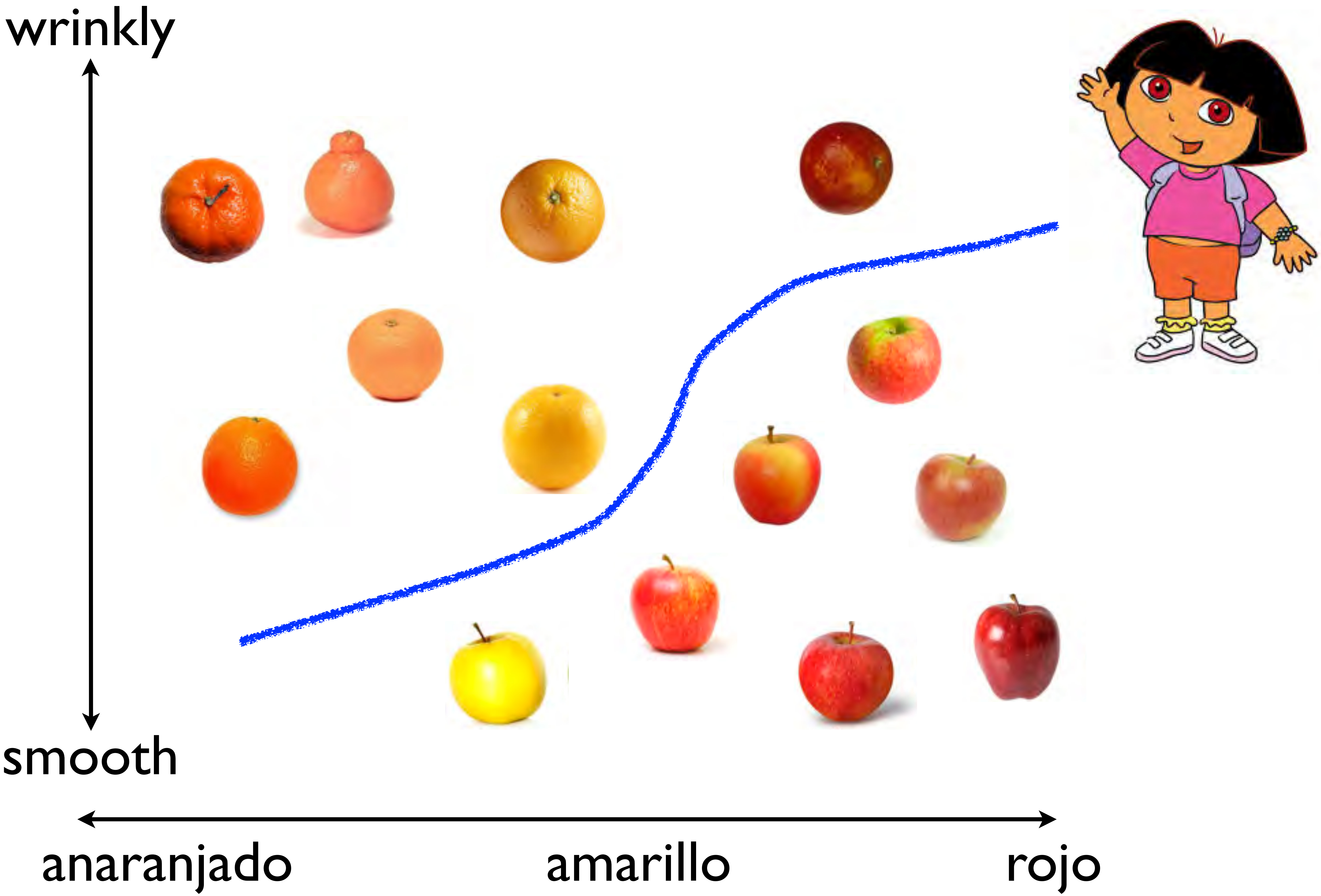
Raw Data

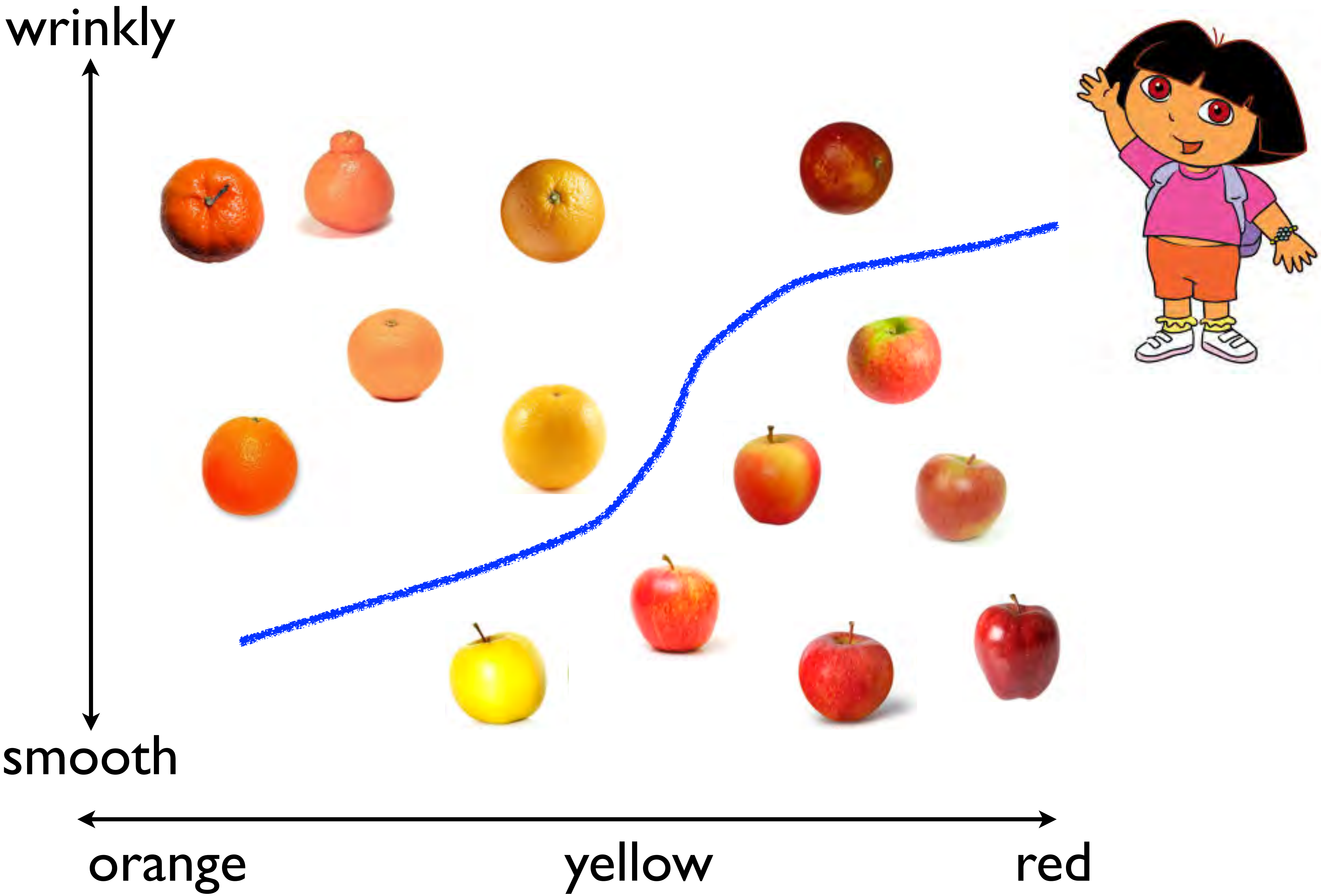


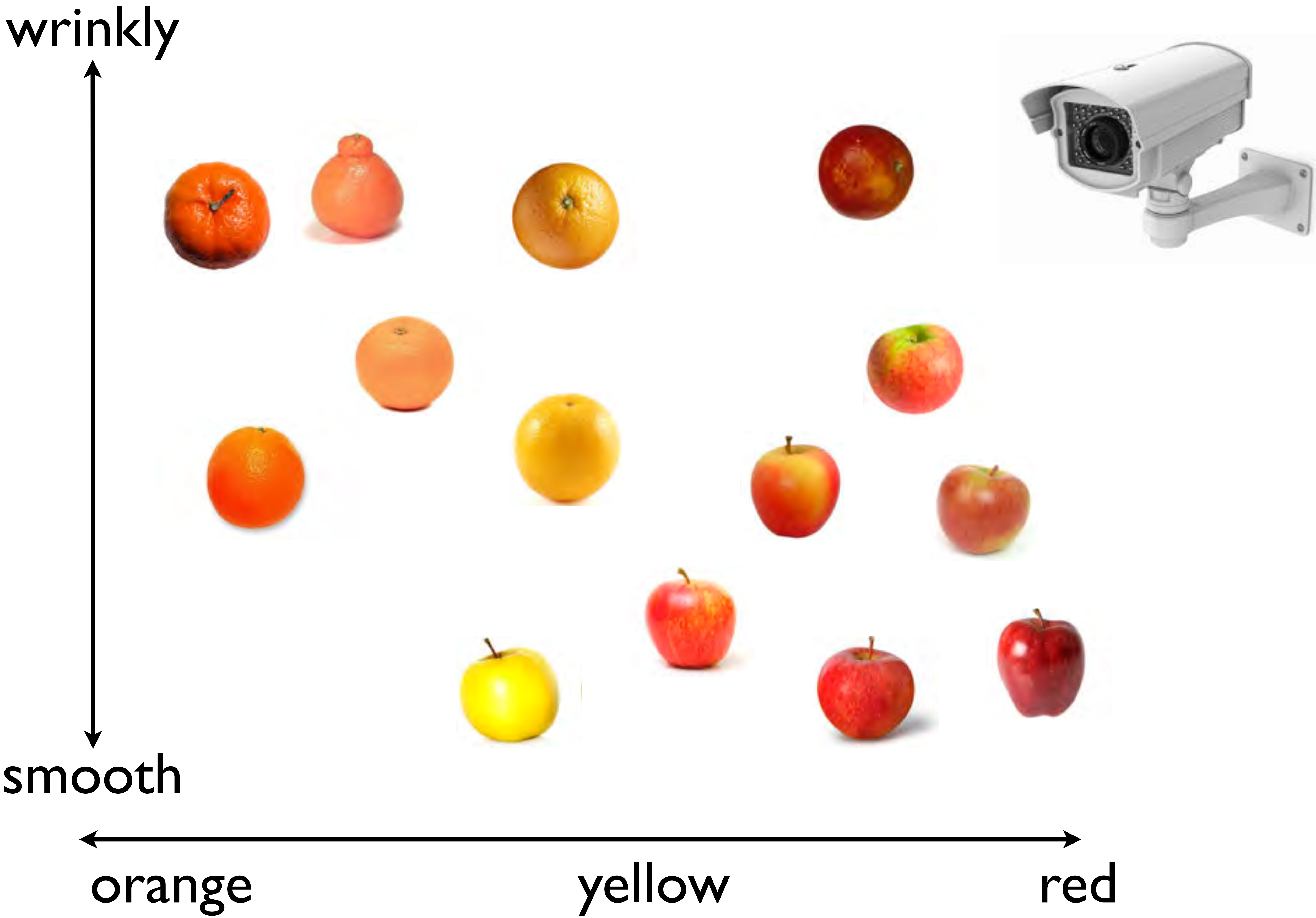
Useful Information

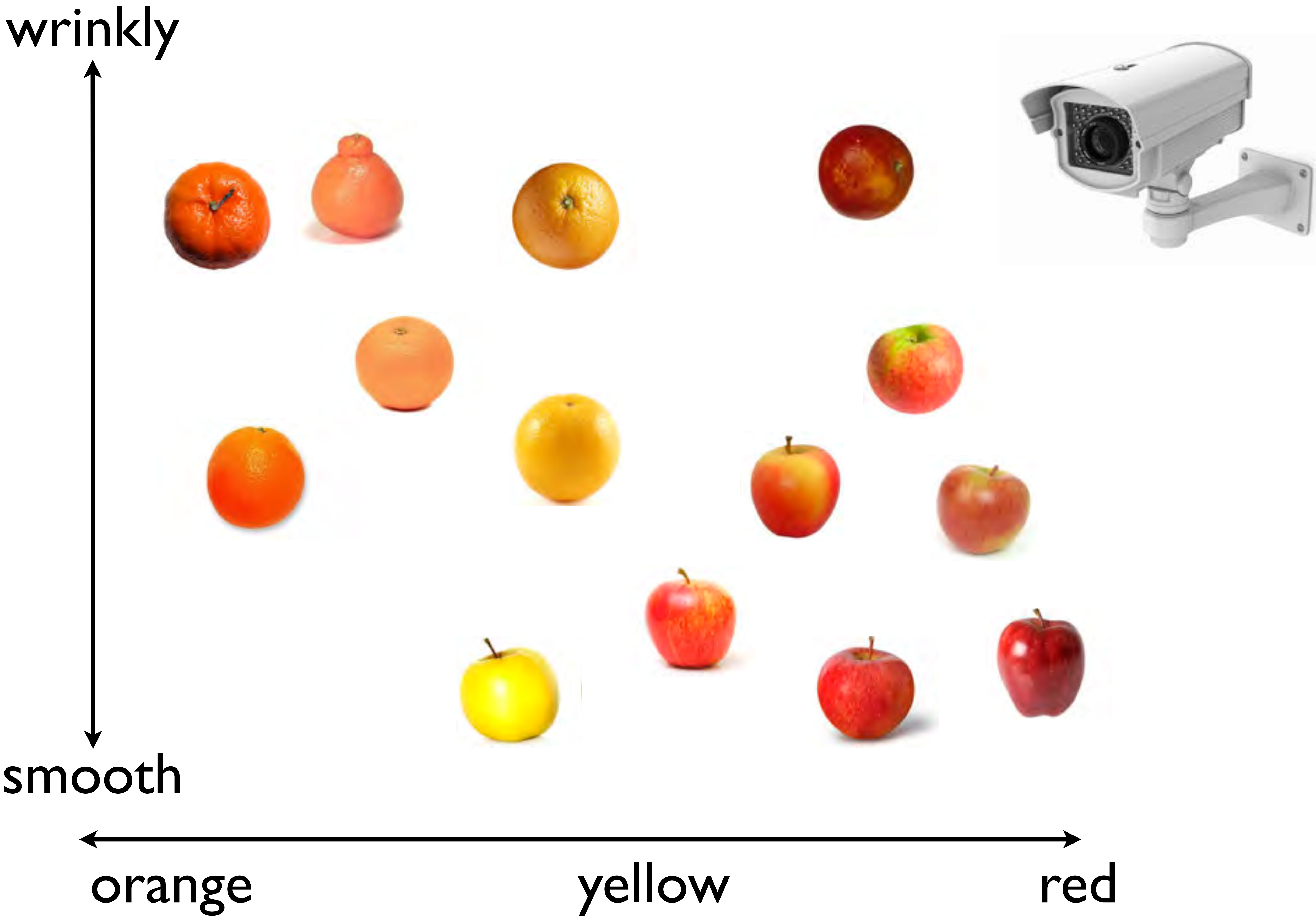

anaranjado
amarillo
rojo

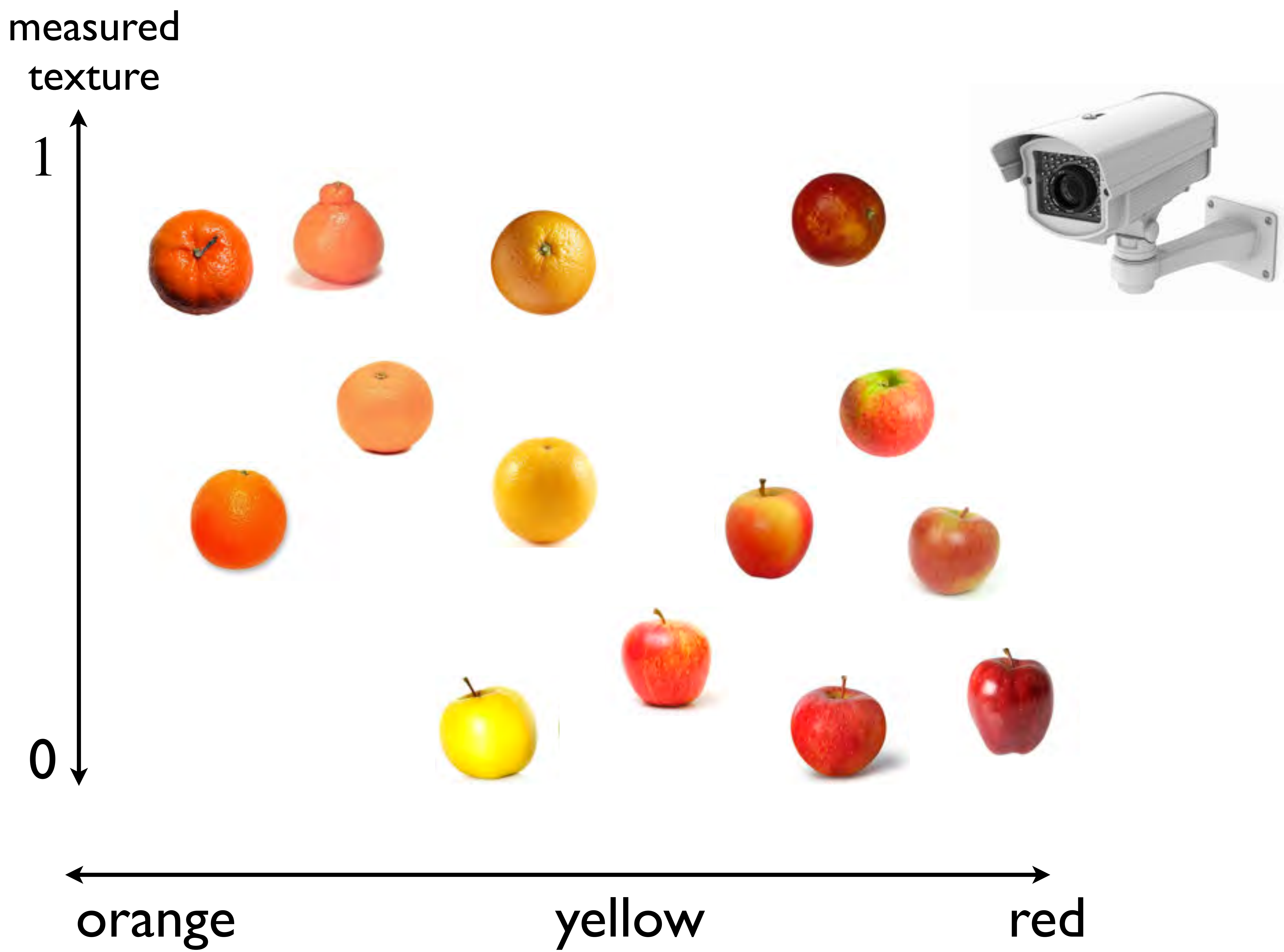

wrinkly
smooth
anaranjado
amarillo
rojo


wrinkly
smooth
anaranjado
amarillo
rojo

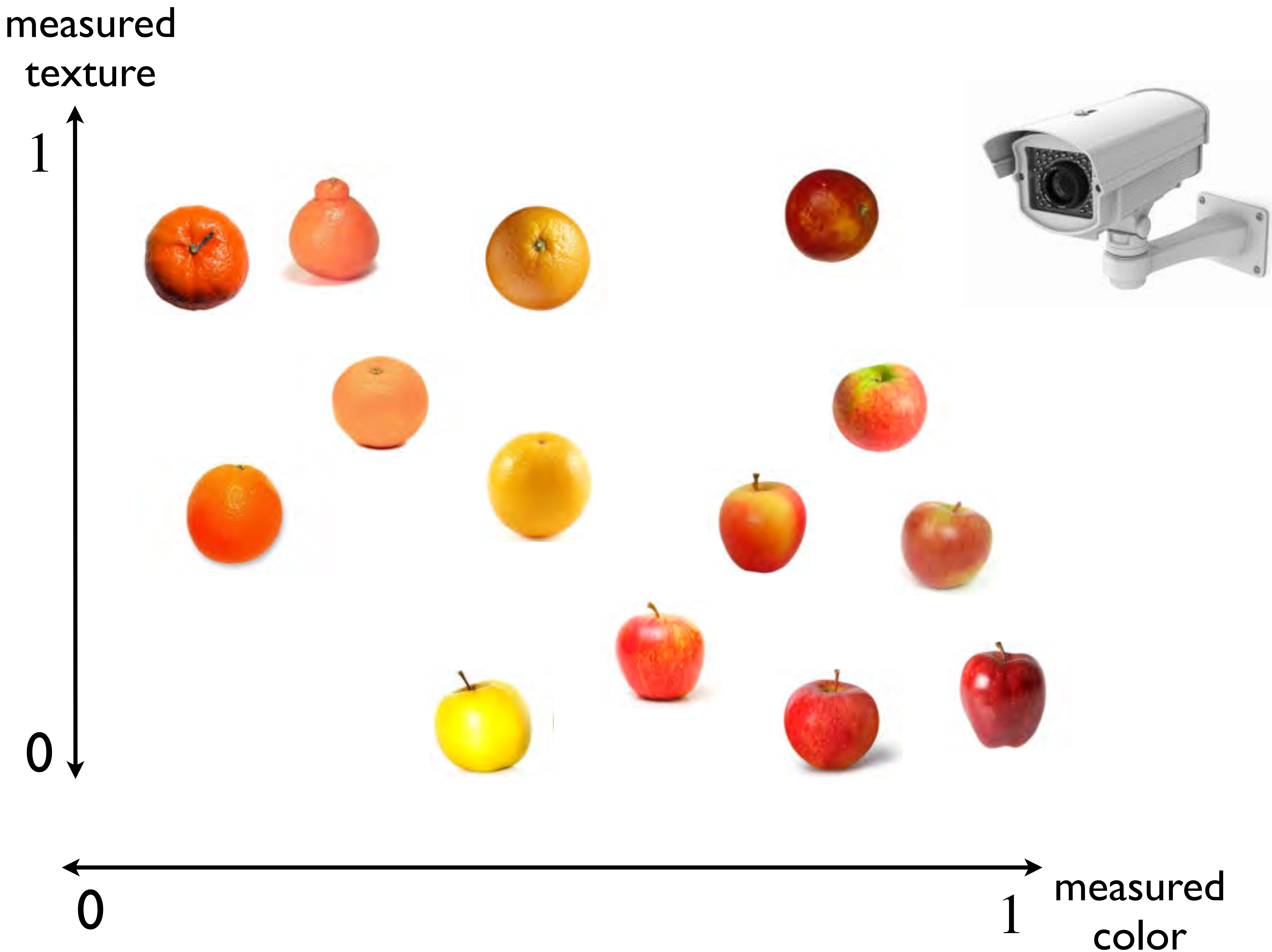

wrinkly
smooth
orange
yellow
red

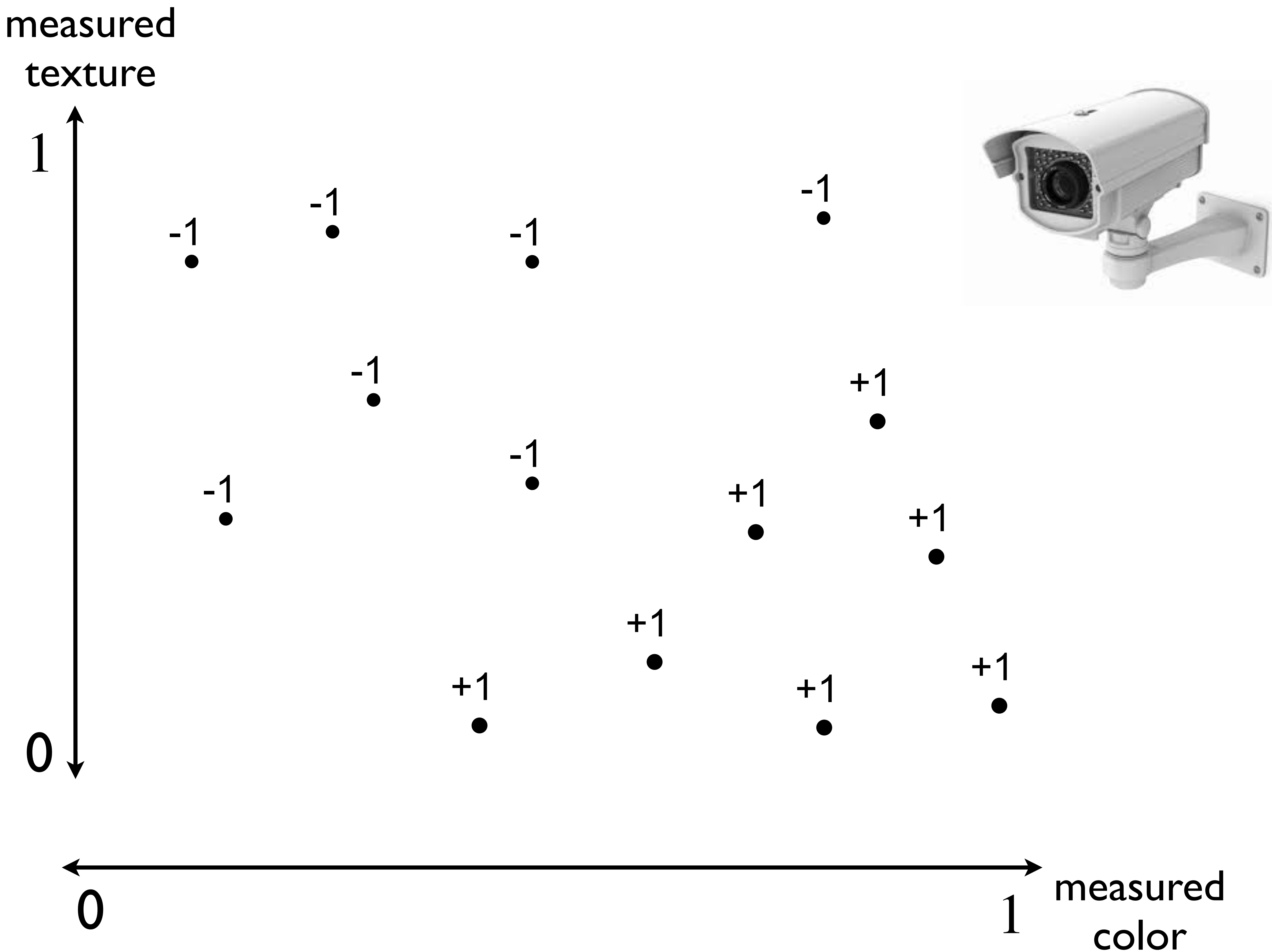

wrinkly
smooth
orange
yellow
red

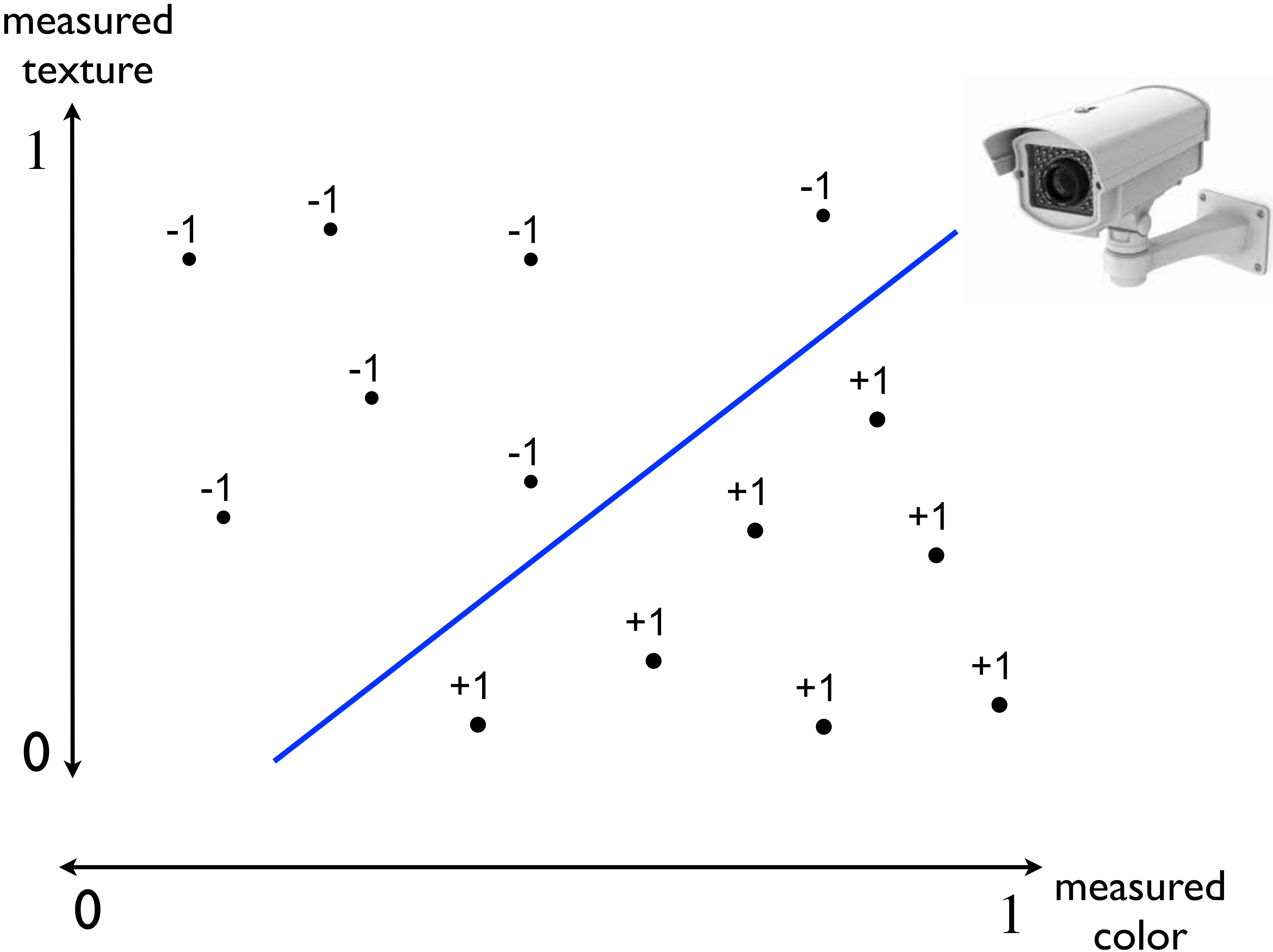

wrinkly
smooth
orange
yellow
red

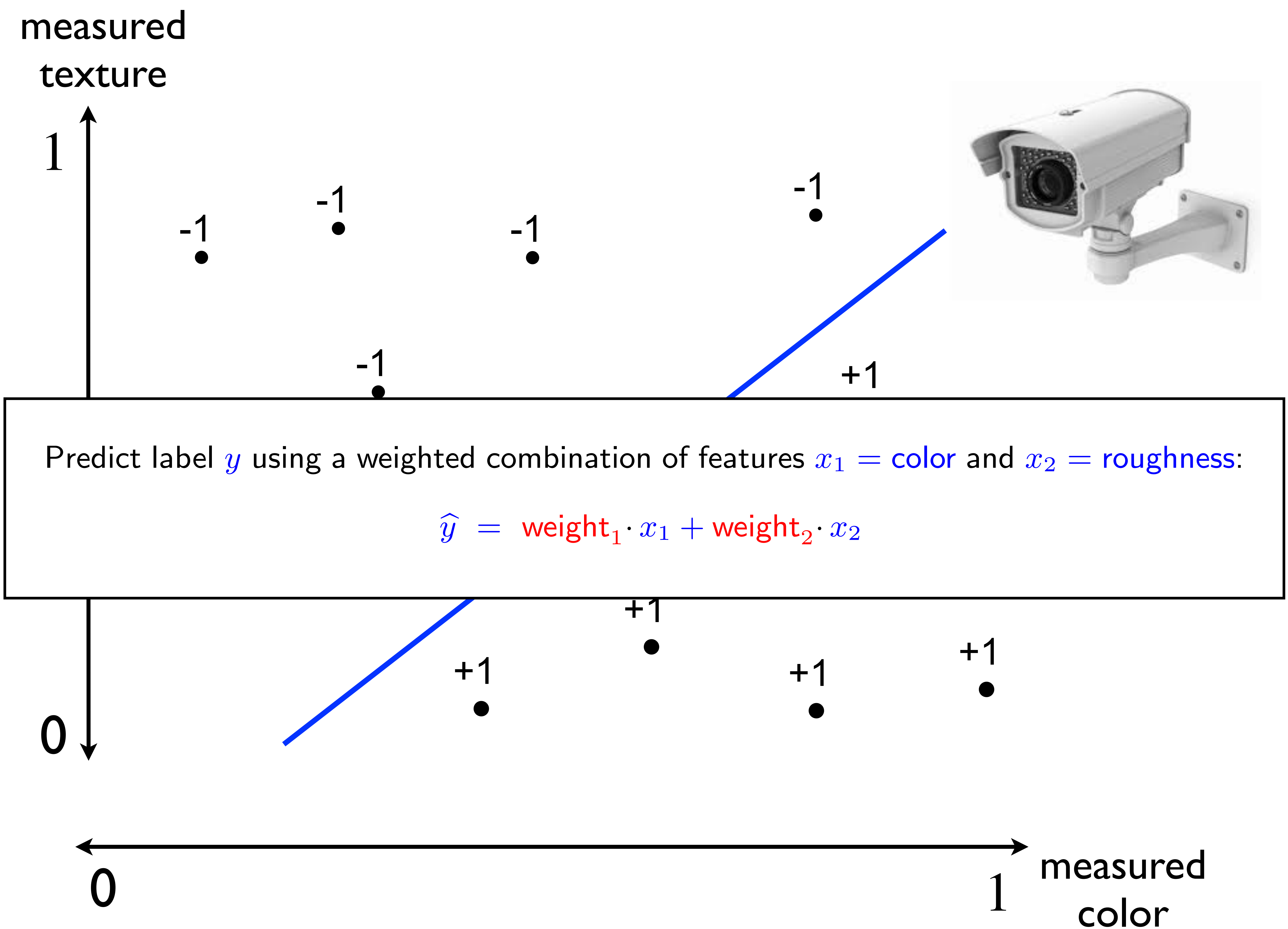


measured
texture
1
0
orange
yellow
red


measured texture
1
0
0
1
measured color


measured
texture
1
0
-1
-1
-1
-1
-1
-1
-1
+1
+1
+1
+1
+1
+1
+1
0
1
measured
color


measured texture
1
0
-1
-1
-1
-1
-1
-1
-1
+1
+1
+1
+1
+1
+1
+1
0
1
measured color



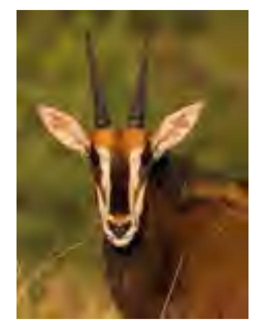
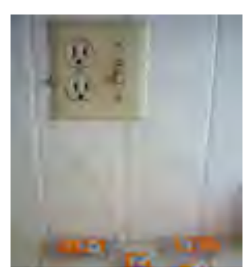
Predict label y using a weighted combination of features x_1 = color and x_2 = roughness:

$$\widehat{y} = \text{weight}_1 \cdot x_1 + \text{weight}_2 \cdot x_2$$

Big Data

IMAGENET

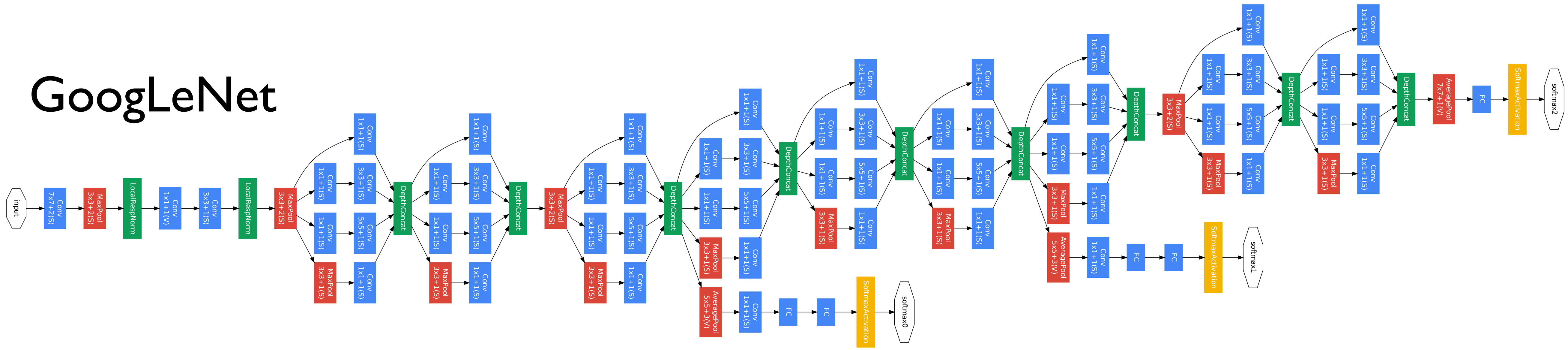
20,000+ concepts
100+ examples of each
14,000,000+ labeled images

accordion	ant	antelope	airplane
			
			

Challenge: Train a machine to recognize all these images

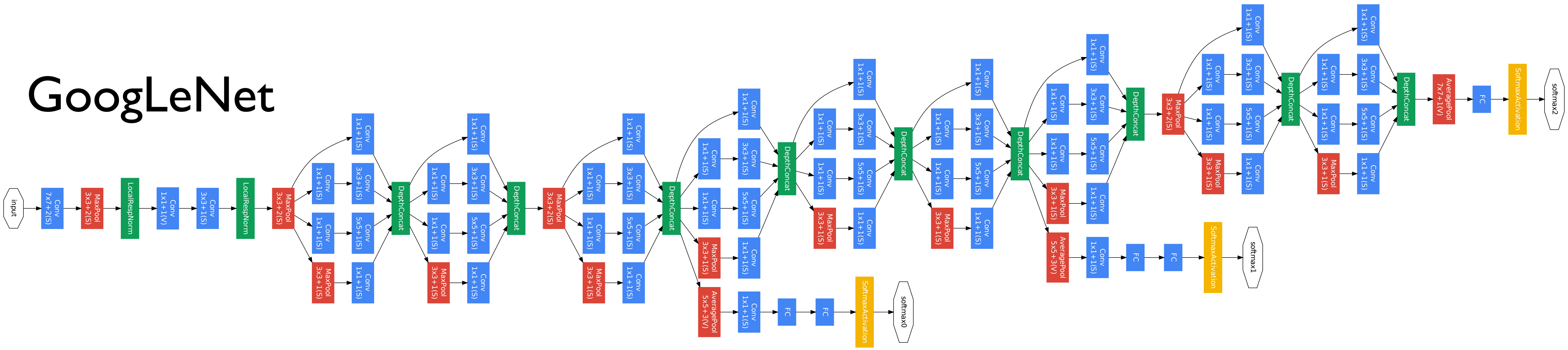
Big Data + Massive Computing = Deep Learning

GoogLeNet



Big Data + Massive Computing = Deep Learning

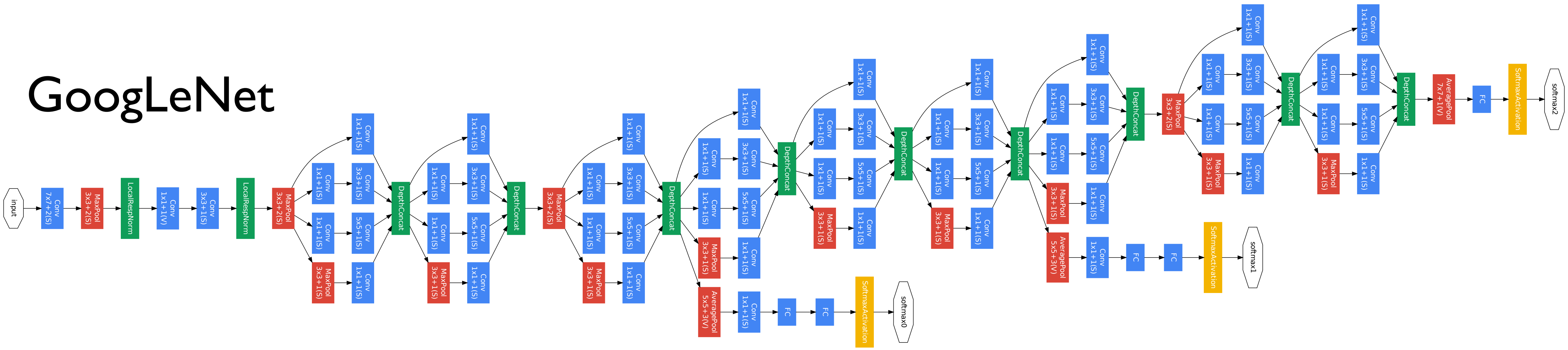
GoogLeNet



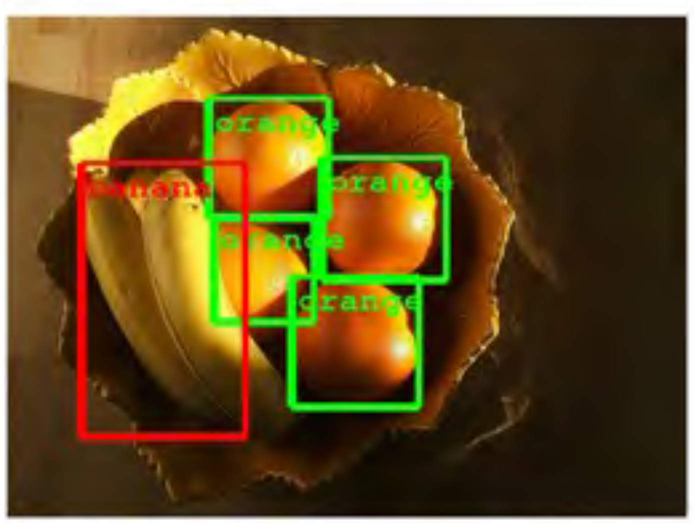
Many layers convolutional filtering + local max.
Each convolutional block has its own **weights**
tuned to maximize accuracy on the training set

Big Data + Massive Computing = Deep Learning

GoogLeNet

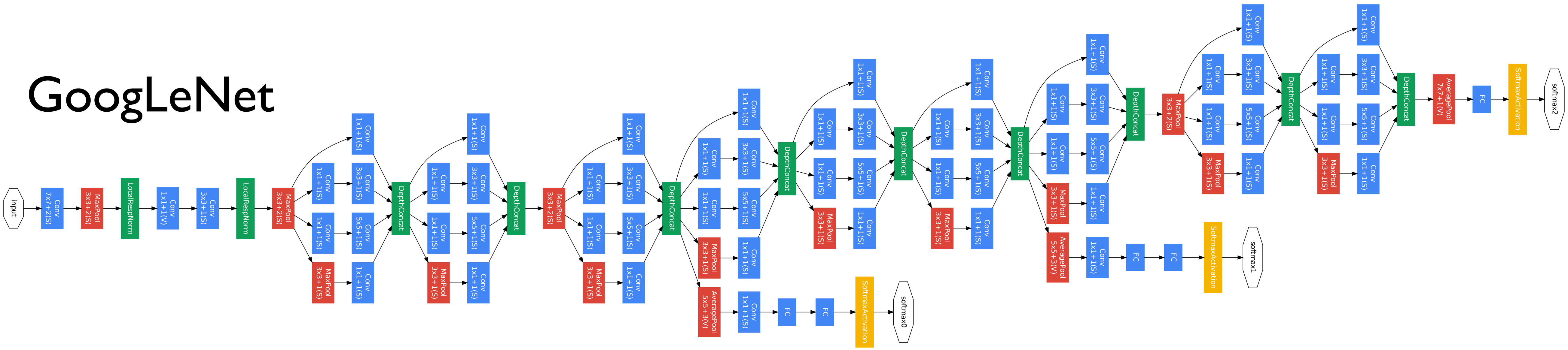


Many layers convolutional filtering + local max.
Each convolutional block has its own **weights**
tuned to maximize accuracy on the training set

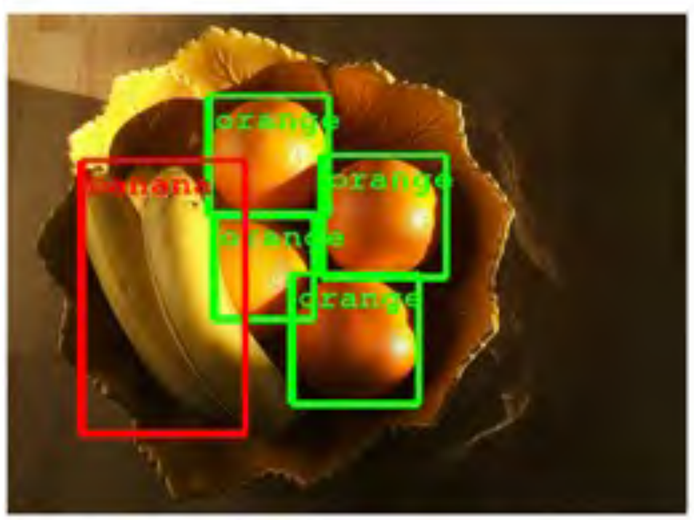


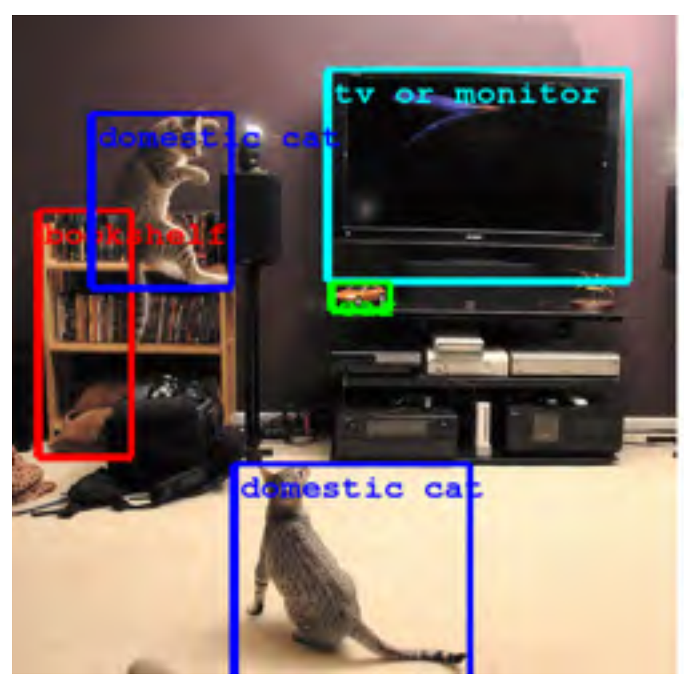
Big Data + Massive Computing = Deep Learning

GoogLeNet



Many layers convolutional filtering + local max.
Each convolutional block has its own **weights**
tuned to maximize accuracy on the training set





Why are neural nets working so well?

- more data
- faster computers

Learning with Thousands of Features

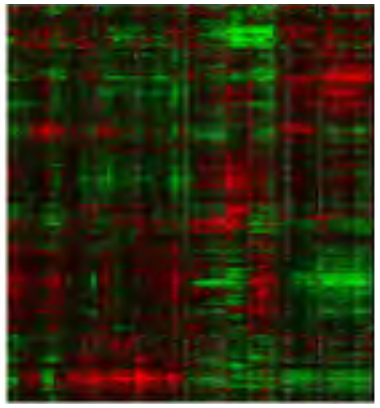
Predict **label** y using a linear combination of **features** $x_1, x_2, \ldots, x_p$:

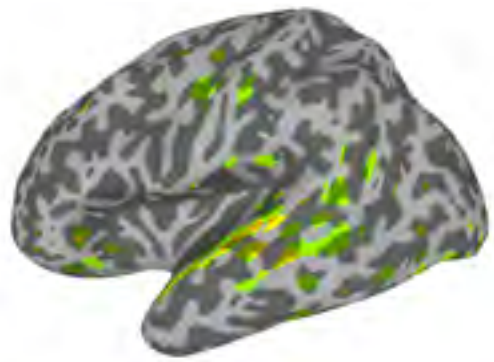
$$\widehat{y} \; = \; w_1 x_1 \; + \; w_2 x_2 \; + \; \ldots \; + \; w_p x_p$$

Learning with Thousands of Features

Predict **label** y using a linear combination of **features** $x_1, x_2, \ldots, x_p$:

$$\widehat{y} = w_1 x_1 + w_2 x_2 + \ldots + w_p x_p$$



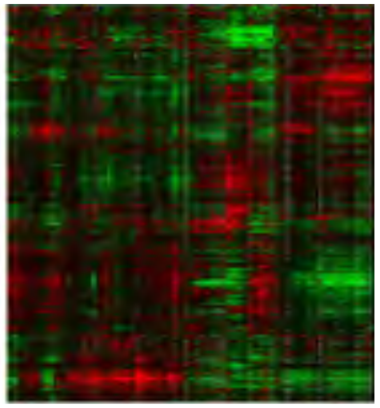


Learning with Thousands of Features

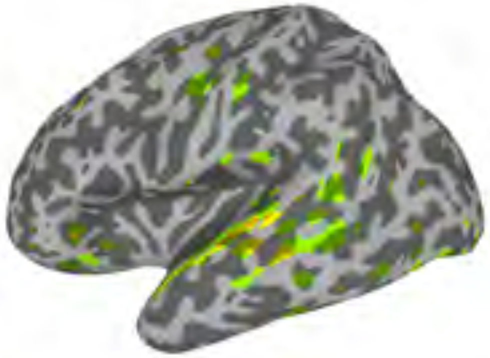
Predict **label** y using a linear combination of **features** $x_1, x_2, \ldots, x_p$:

$$\widehat{y} \; = \; w_1 x_1 \; + \; w_2 x_2 \; + \; \ldots \; + \; w_p x_p$$

use n training examples to find best weights



$$\begin{aligned} y_1 &\approx w_1 x_{11} + w_2 x_{12} + \cdots + w_p x_{1p} \\ y_2 &\approx w_1 x_{21} + w_2 x_{22} + \cdots + w_p x_{2p} \\ \vdots & \qquad\qquad \vdots \\ y_n &\approx w_1 x_{n1} + w_2 x_{n2} + \cdots + w_p x_{np} \end{aligned}$$

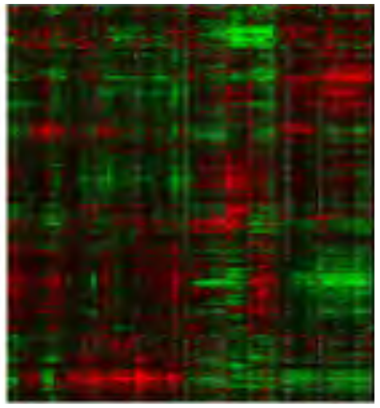


Learning with Thousands of Features

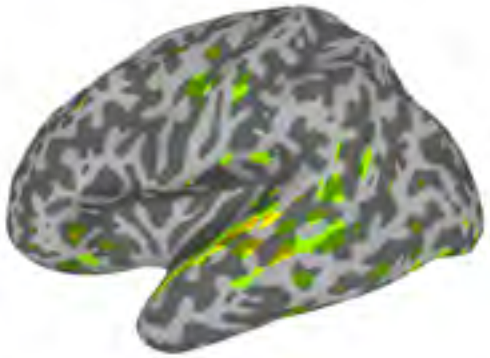
Predict **label** y using a linear combination of **features** $x_1, x_2, \ldots, x_p$:

$$\widehat{y} = w_1 x_1 + w_2 x_2 + \ldots + w_p x_p$$

use n training examples to find best weights



$$\begin{aligned} y_1 &\approx w_1 x_{11} + w_2 x_{12} + \cdots + w_p x_{1p} \\ y_2 &\approx w_1 x_{21} + w_2 x_{22} + \cdots + w_p x_{2p} \\ &\vdots \\ y_n &\approx w_1 x_{n1} + w_2 x_{n2} + \cdots + w_p x_{np} \end{aligned}$$



In many applications

- $p \gg n \Rightarrow$ fewer equations than unknowns!

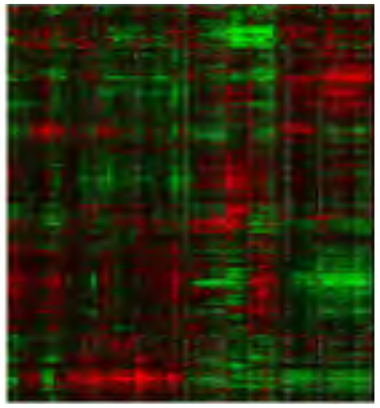
Learning with Thousands of Features

Predict **label** y using a linear combination of **features** $x_1, x_2, \ldots, x_p$:

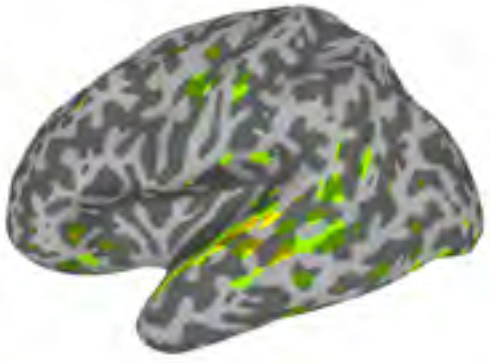
$$\widehat{y} = w_1 x_1 + w_2 x_2 + \ldots + w_p x_p$$

use n training examples to find best weights

$$\begin{aligned} y_1 &\approx w_1 x_{11} + w_2 x_{12} + \cdots + w_p x_{1p} \\ y_2 &\approx w_1 x_{21} + w_2 x_{22} + \cdots + w_p x_{2p} \\ &\vdots \qquad\qquad \vdots \\ y_n &\approx w_1 x_{n1} + w_2 x_{n2} + \cdots + w_p x_{np} \end{aligned}$$



only a few genes important in disease



only a few active brain areas

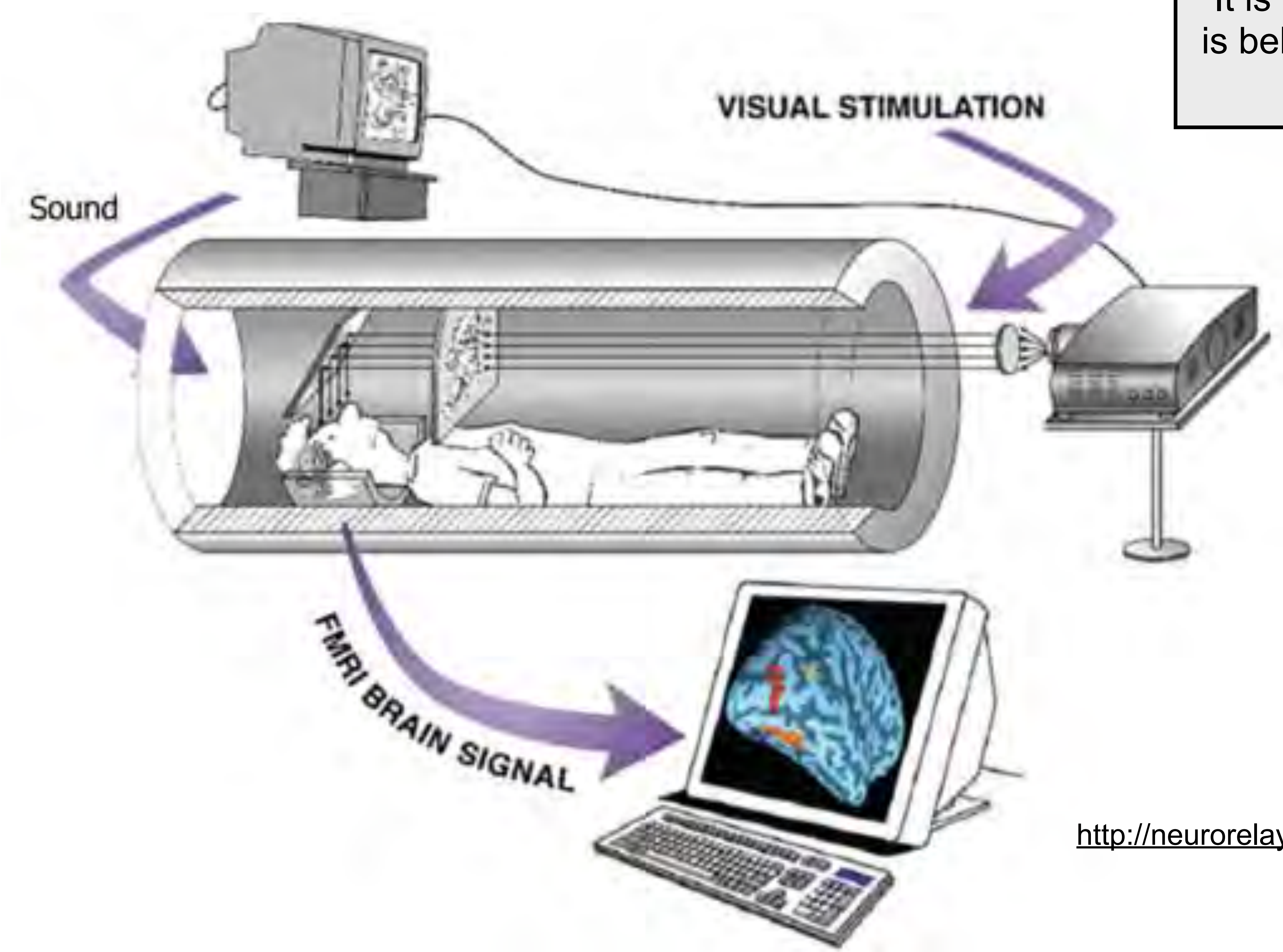
In many applications

- $p \gg n \Rightarrow$ fewer equations than unknowns!
- many of the weights $w_1, w_2, \ldots, w_p$ should be zero

Machines Reading Minds with fMRI

sentence: +1

"It is true that the star is below the plus."

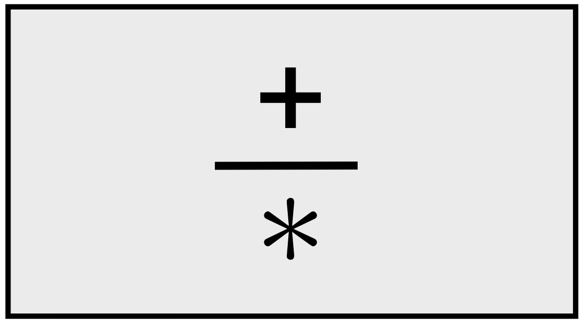


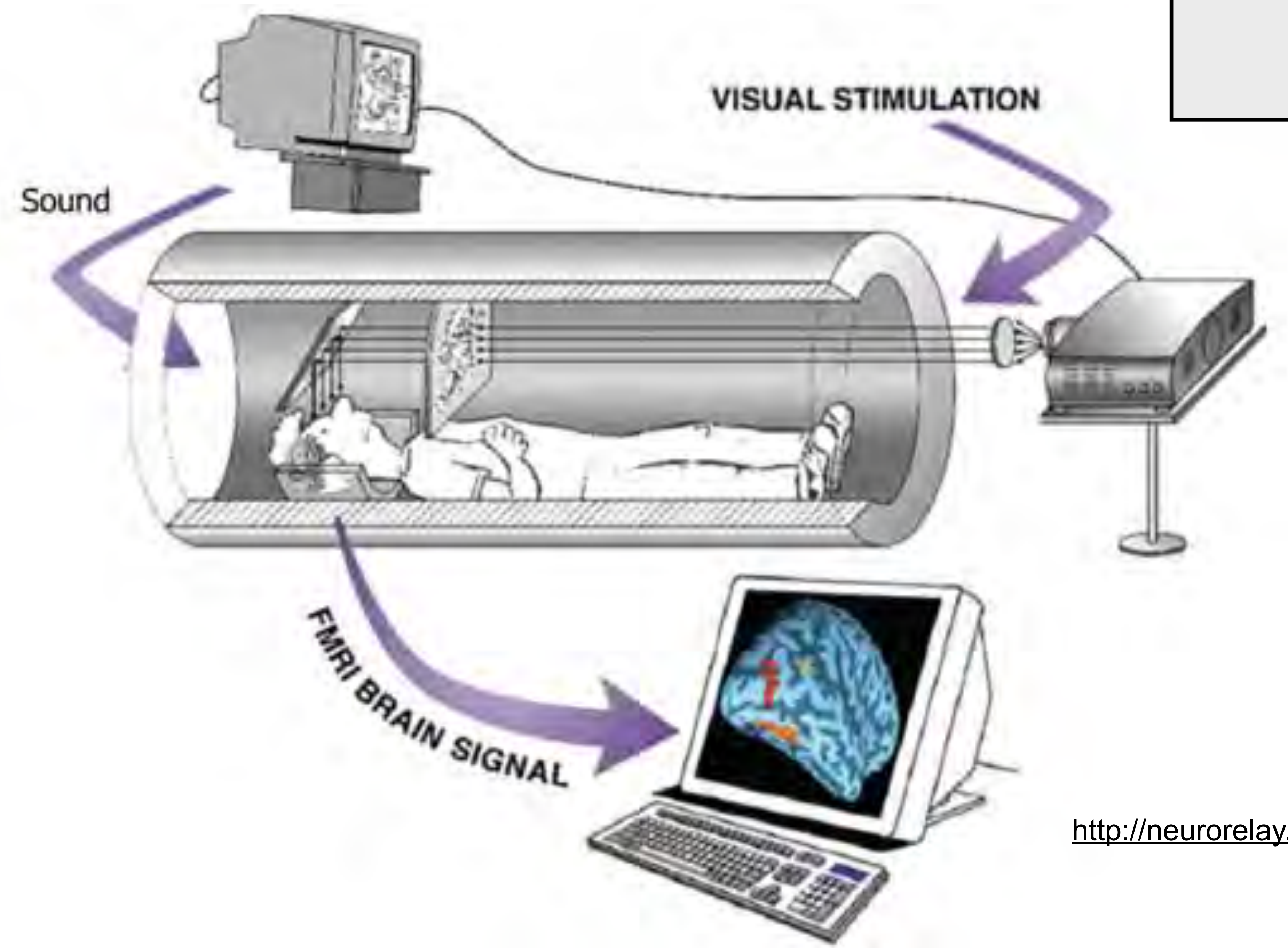
http://neurorelay.com

Wang, Mitchell, Hutchinson '03

Machines Reading Minds with fMRI

image: -1

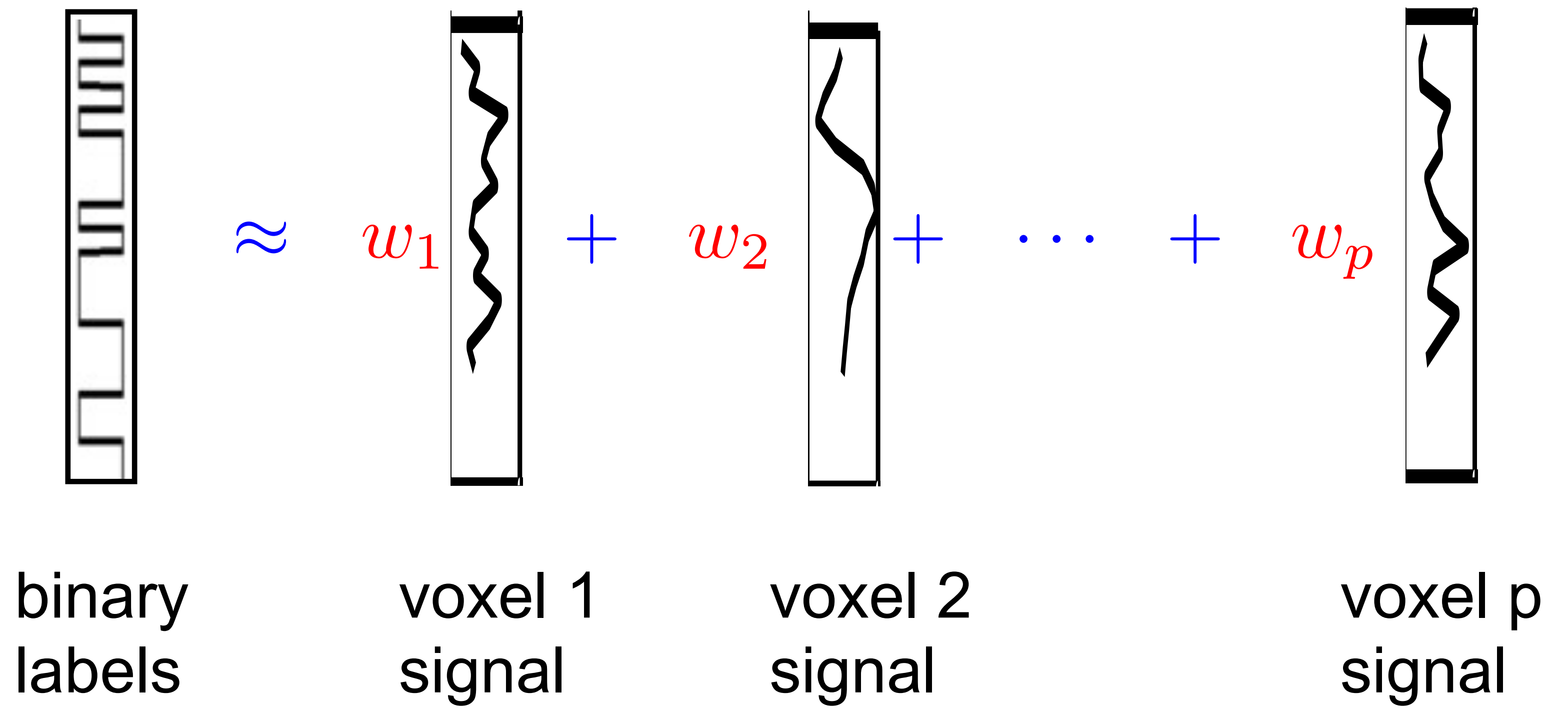




http://neurorelay.com

Wang, Mitchell, Hutchinson '03

Predicting Stimulus from fMRI Signals



Challenge: Train a machine to predict label (picture: -1 or sentence: +1)

Mapping Brain Activity via Optimization

$$\widehat{\boldsymbol{w}} \;=\; \arg\min_{\boldsymbol{w}} \; \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} w_j x_{ij} \right)^2 \;+\; \lambda \sum_{j=1}^{p} |w_j| \right\}$$

Rao, Cox, Rogers, and Nowak ’13

Mapping Brain Activity via Optimization

$$\widehat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}} \left\{ \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} w_j x_{ij}\right)^2 + \lambda \sum_{j=1}^{p} |w_j| \right\}$$

fit to data

in a few cortical regions

Rao, Cox, Rogers, and Nowak '13

Mapping Brain Activity via Optimization

$$\widehat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} w_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} |w_j| \right\}$$

fit to data

in a few cortical regions

prediction accuracy 70%-75% across multiple subjects

Rao, Cox, Rogers, and Nowak '13

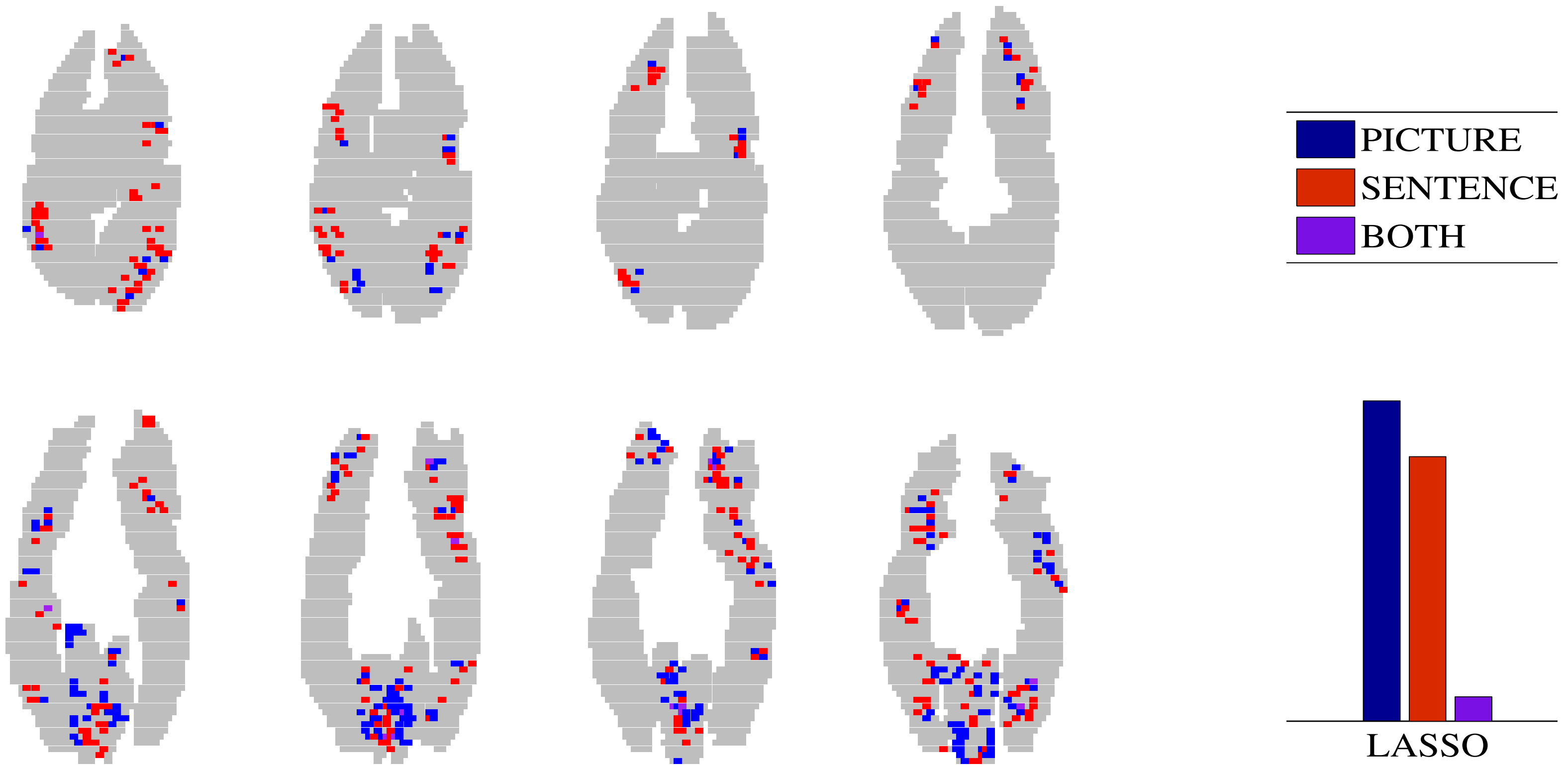
Mapping Brain Activity via Optimization

$$\widehat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}} \left\{ \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p} w_j x_{ij}\right)^2 + \lambda \sum_{j=1}^{p} |w_j| \right\}$$

fit to data

in a few
cortical
regions

prediction accuracy 70%-75%
across multiple subjects



Rao, Cox, Rogers, and Nowak '13

People + Machines

- how do humans reason about complicated concepts?
- what are people's preferences?
- how can people train machines with minimal human effort?

hipster bartender

hipster bartender

Bartender: “What beer would you like?”

hipster bartender

Bartender: “What beer would you like?”

AI: “Hmm... I prefer red wine”

hipster bartender

Bartender: “What beer would you like?”

AI: “Hmm... I prefer red wine”

Bartender: “Try these two samples. Do you prefer A or B?”

hipster bartender

Bartender: “What beer would you like?”

AI: “Hmm... I prefer red wine”

Bartender: “Try these two samples. Do you prefer A or B?”

AI: “B”

hipster bartender

Bartender: “What beer would you like?”

AI: “Hmm... I prefer red wine”

Bartender: “Try these two samples. Do you prefer A or B?”

AI: “B”

Bartender: “Ok try these two: C or D?” ....

robot bartender

Bartender: “What beer would you like?”

AI: “Hmm... I prefer red wine”

Bartender: “Try these two samples. Do you prefer A or B?”

AI: “B”

Bartender: “Ok try these two: C or D?” ....

Beer Maps

- can we train a machine to learn a beer map?
- can the machine use a map to recommend beers?

ratebeer

Two Hearted Ale - Input ~2500 reviews

http://www.ratebeer.com/beer/two-hearted-ale/

3.8 AROMA 8/10 APPEARANCE 4/5 TASTE 8/10 PALATE 3/5 OVERALL 15/20

fonefan (25678) - VestJylland, DENMARK - JAN 18, 2009

Bottle 355ml.

Clear light to medium yellow orange color with a average, frothy, good lacing, fully lasting, off-white head. Aroma is moderate to heavy malty, moderate to heavy hoppy, perfume, grapefruit, orange shell, soap. Flavor is moderate to heavy sweet and bitter with a average to long duration. Body is medium, texture is oily, carbonation is soft. [250908]

4 AROMA 8/10 APPEARANCE 4/5 TASTE 7/10 PALATE 4/5 OVERALL 17/20

Ungstrup (24358) - Oamaru, NEW ZEALAND - MAR 31, 2005

An orange beer with a huge off-white head. The aroma is sweet and very freshly hoppy with notes of hop oils - very powerful aroma. The flavor is sweet and quite hoppy, that gives flavors of oranges, flowers as well as hints of grapefruit. Very refreshing yet with a powerful body.

---

ratebeer

Two Hearted Ale - Input ~2500 reviews

http://www.ratebeer.com/beer/two-hearted-ale/

3.8 AROMA 8/10 APPEARANCE 4/5 TASTE 8/10 PALATE 3/5 OVERALL 15/20

fonefan (25678) - VestJylland, DENMARK - JAN 18, 2009

Bottle 355ml.

Clear light to medium yellow orange color with a average, frothy, good lacing, fully lasting, off-white head. Aroma is moderate to heavy malty, moderate to heavy hoppy, perfume, grapefruit, orange shell, soap. Flavor is moderate to heavy sweet and bitter with a average to long duration. Body is medium, texture is oily, carbonation is soft. [250908]

4 AROMA 8/10 APPEARANCE 4/5 TASTE 7/10 PALATE 4/5 OVERALL 17/20

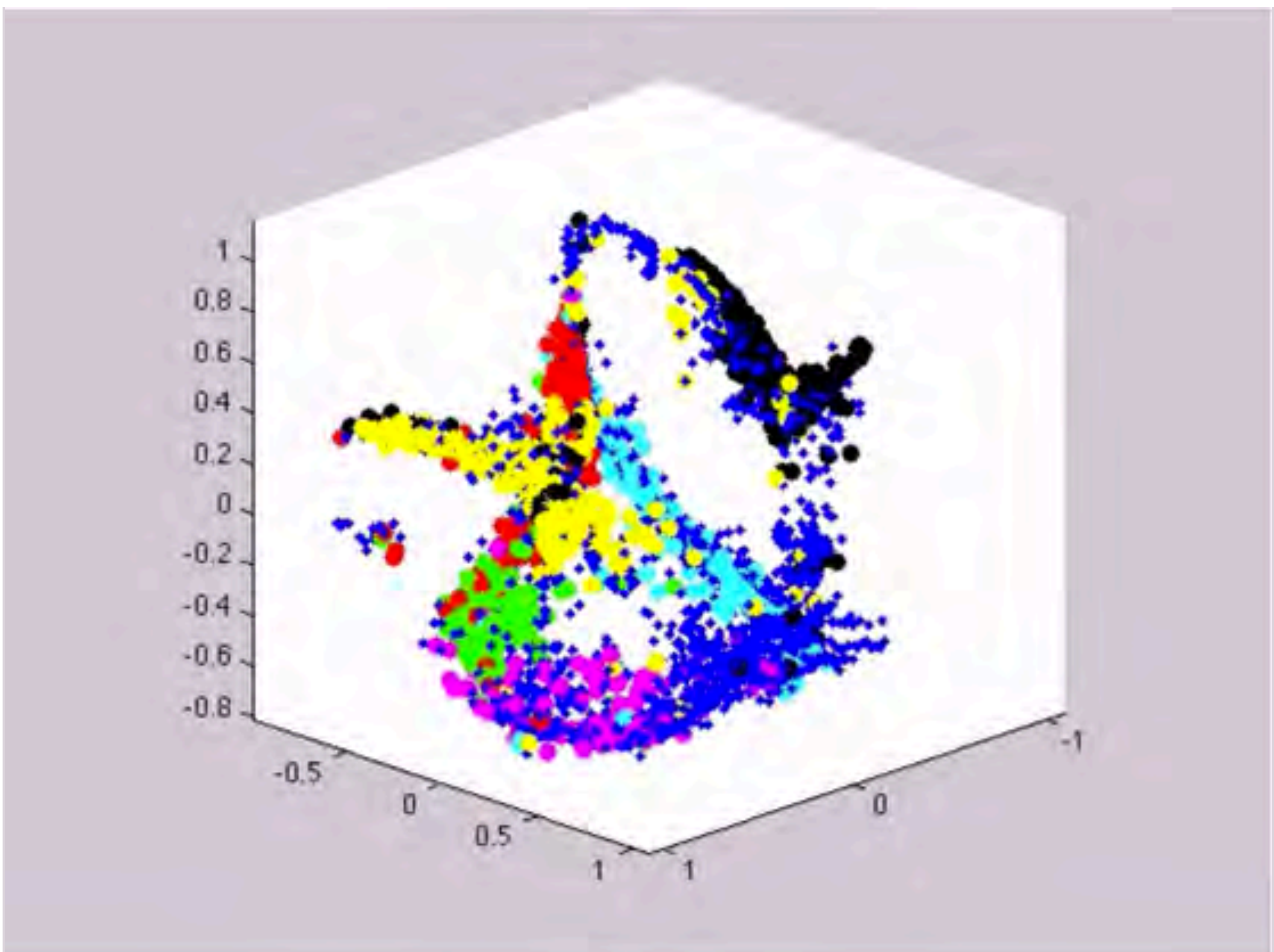
Ungstru

An orange beer with a

very powerful aroma.

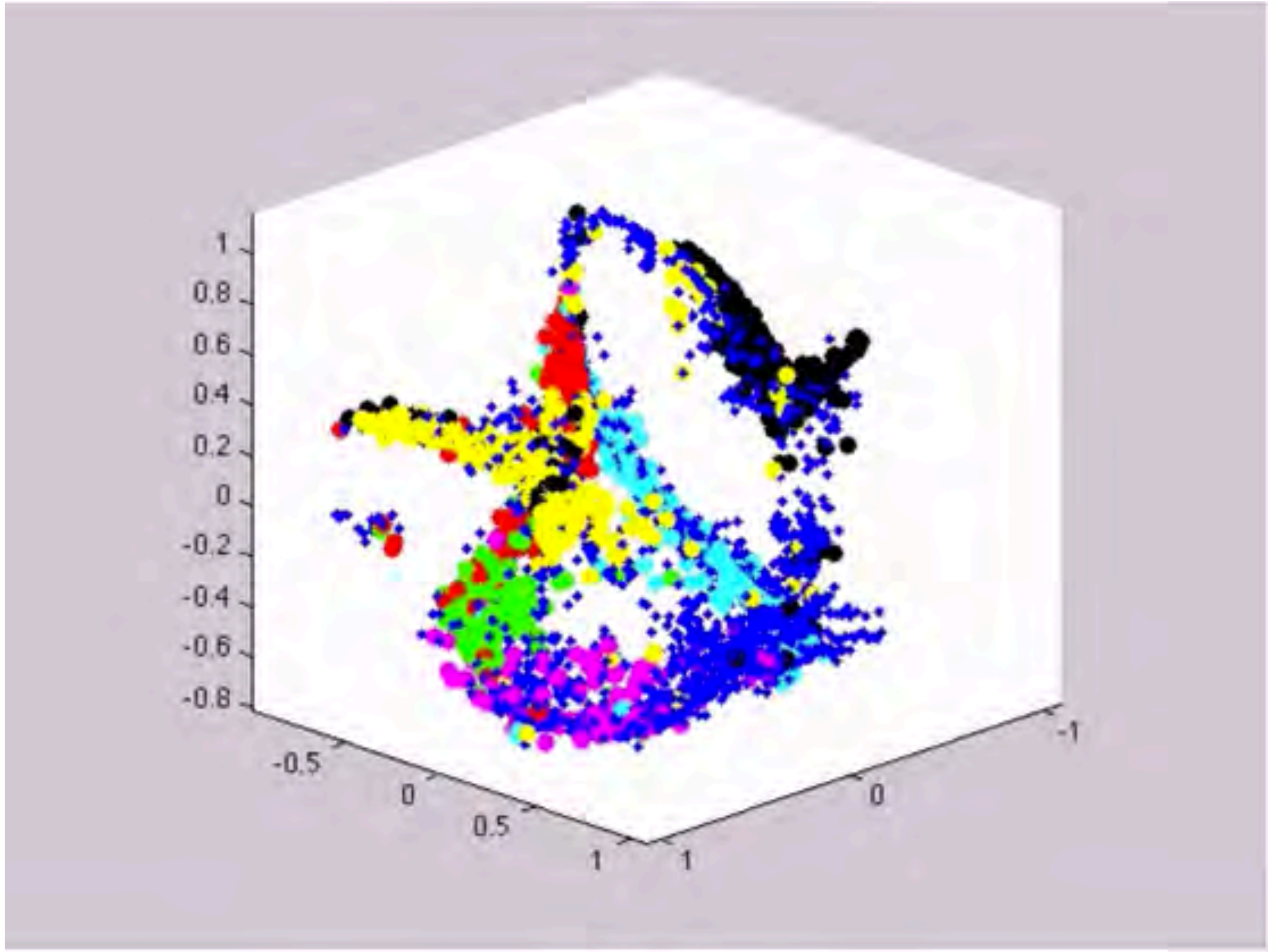
of grapefruit. Very refr

Map (cluster) beers based on word cloud similarities



Red = IPA
Green = Pale Ale
Magenta = Amber Ale
Cyan = Lager + Pilsener
Yellow = Belgians (light + dark)
Black = Stout + Porter
Blue = Everything else

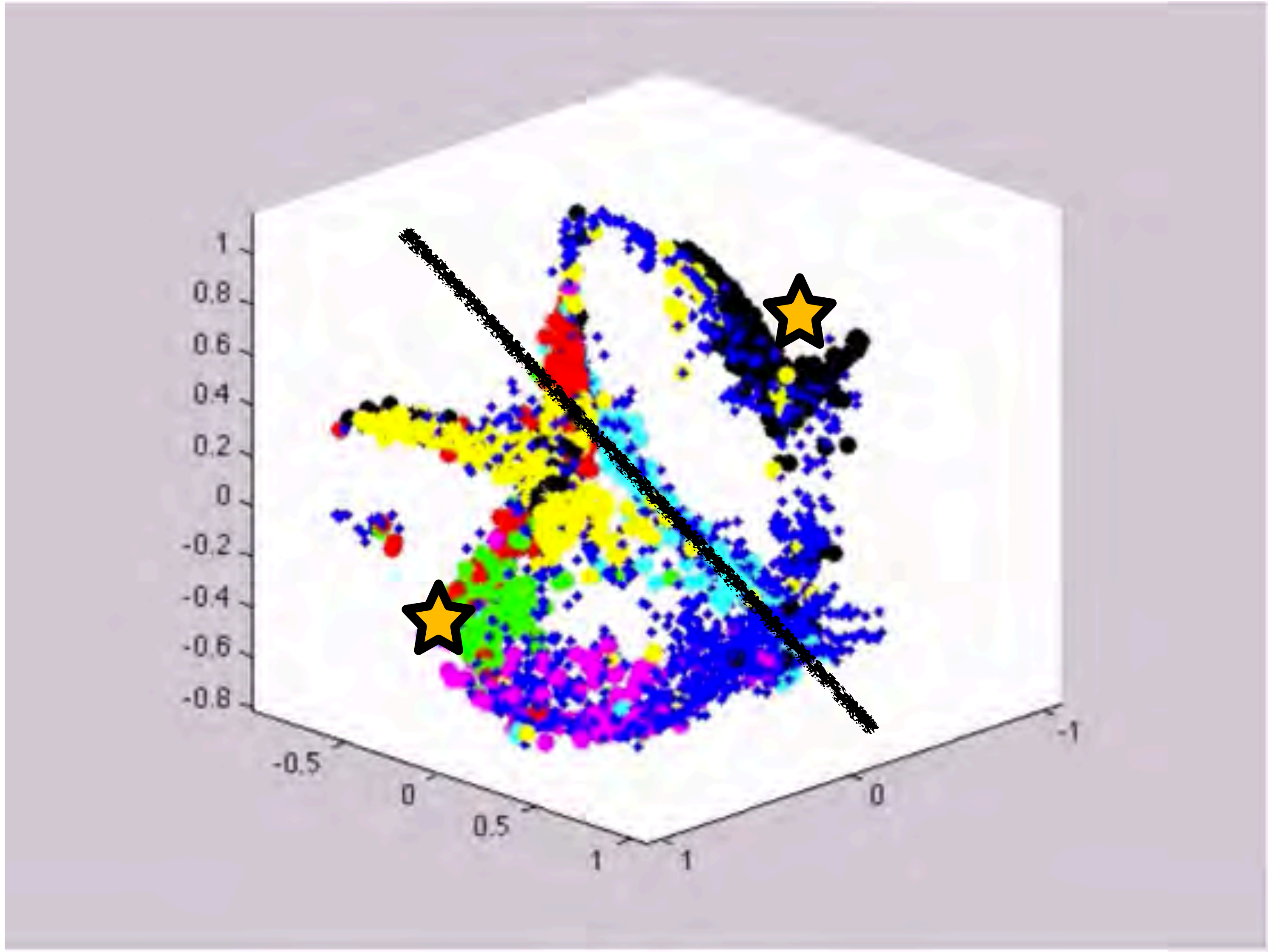
Ask AI to compare or rate strategically selected beers



Red = IPA
Green = Pale Ale
Magenta = Amber Ale
Cyan = Lager + Pilsener
Yellow = Belgians (light + dark)
Black = Stout + Porter
Blue = Everything else

Jamieson and Nowak '11

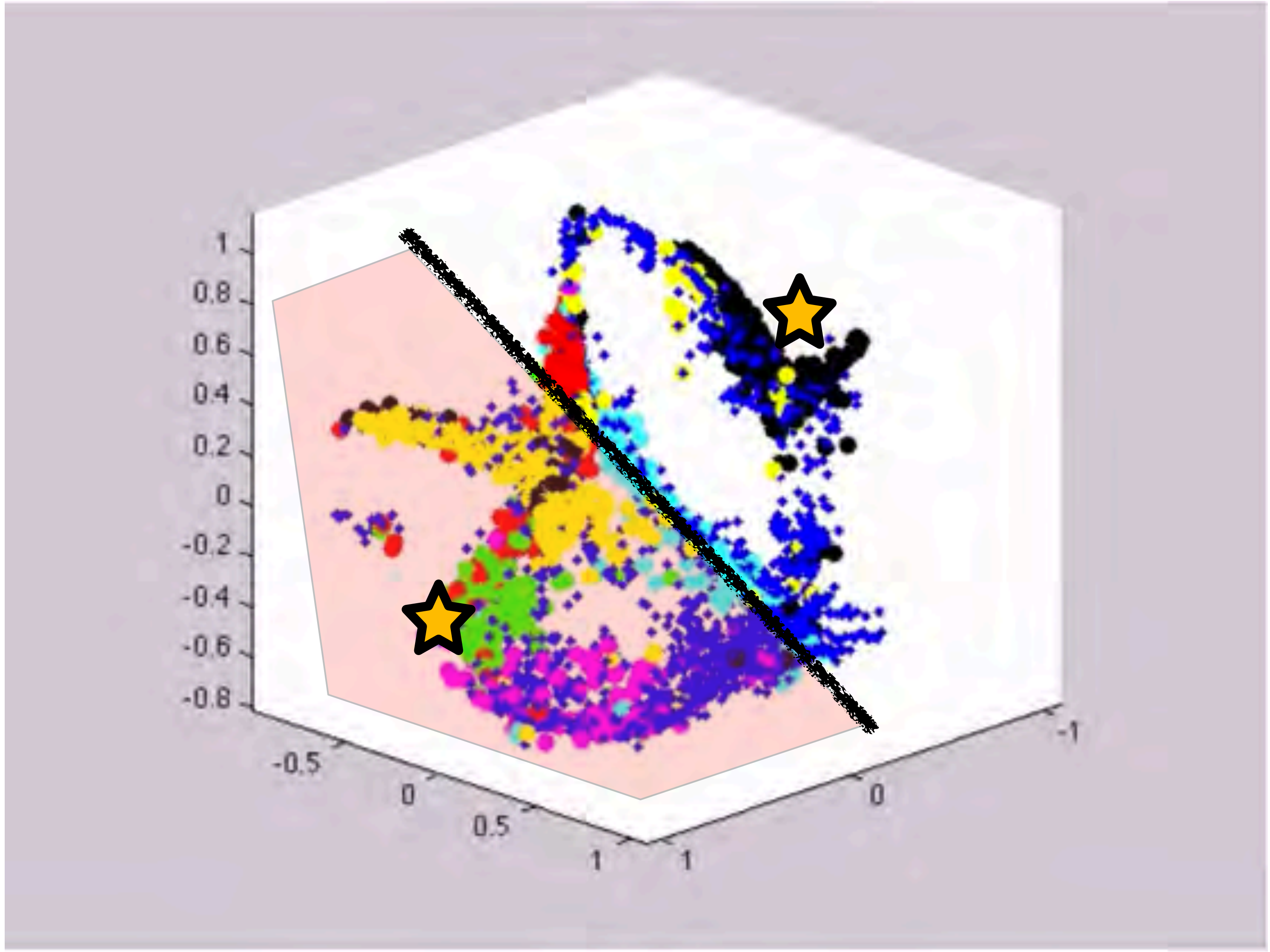
Ask AI to compare or rate strategically selected beers



Red = IPA
Green = Pale Ale
Magenta = Amber Ale
Cyan = Lager + Pilsener
Yellow = Belgians (light + dark)
Black = Stout + Porter
Blue = Everything else

Jamieson and Nowak '11

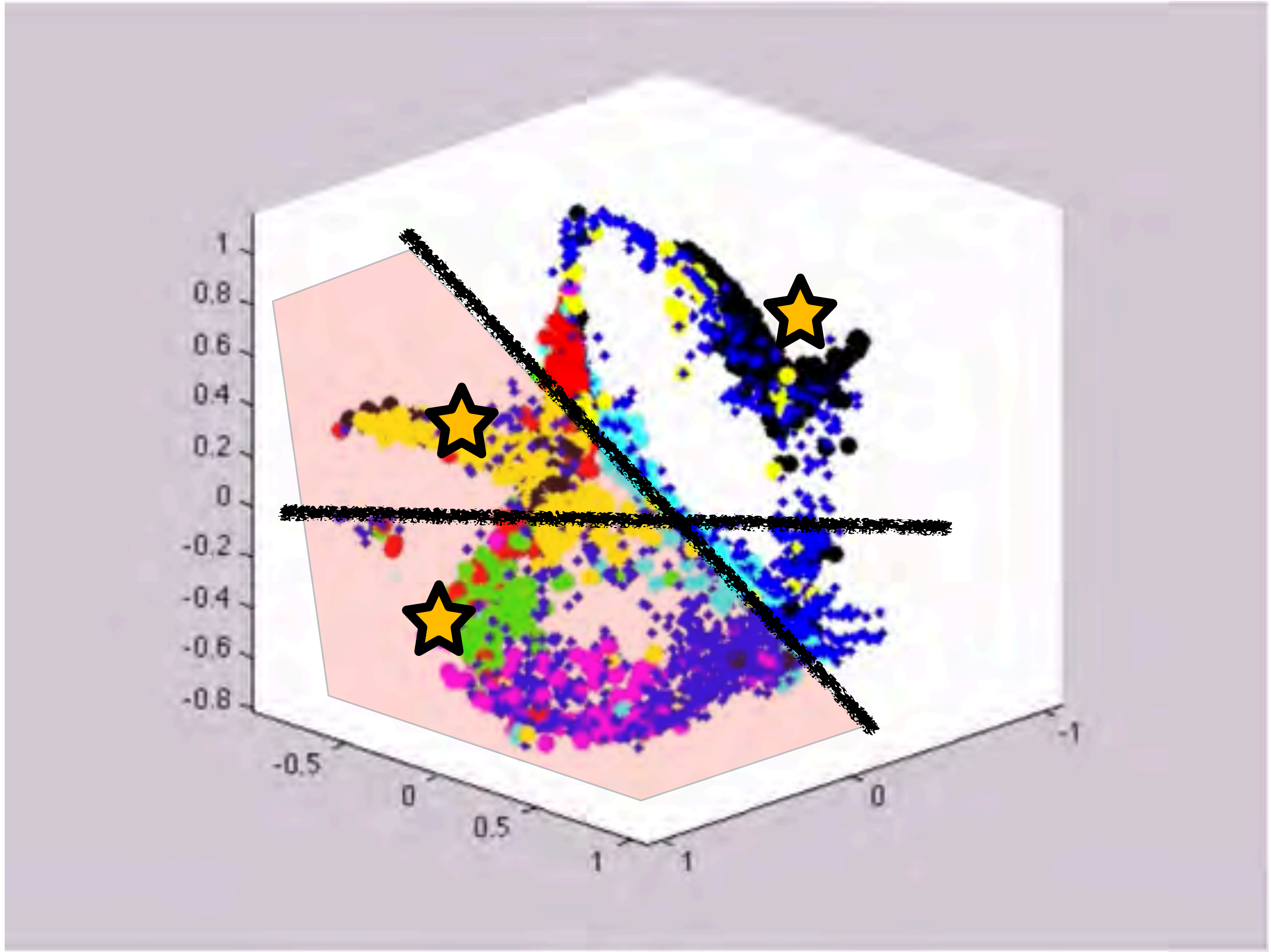
Ask AI to compare or rate strategically selected beers



Red = IPA
Green = Pale Ale
Magenta = Amber Ale
Cyan = Lager + Pilsener
Yellow = Belgians (light + dark)
Black = Stout + Porter
Blue = Everything else

Jamieson and Nowak ’11

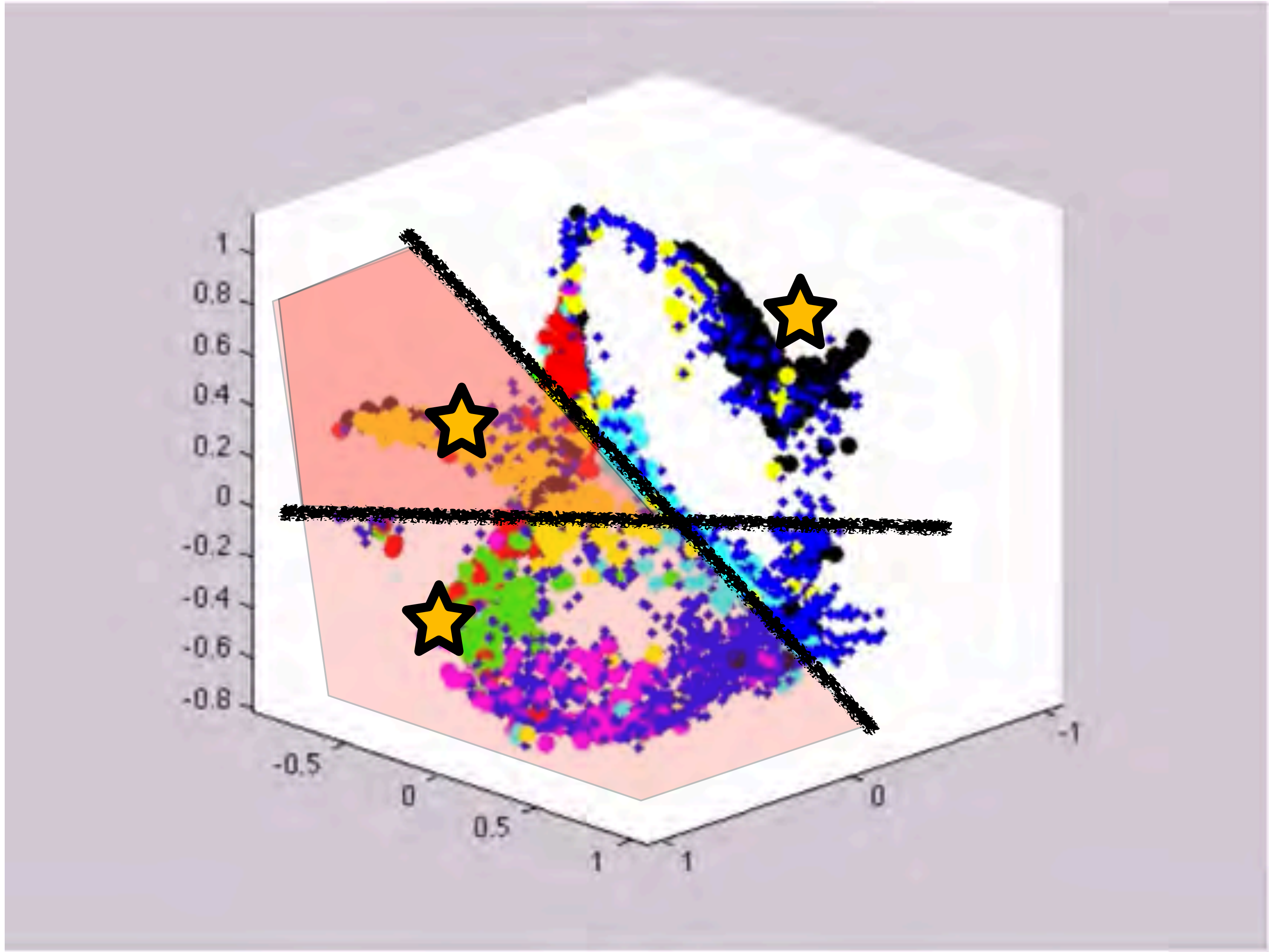
Ask AI to compare or rate strategically selected beers



Red = IPA
Green = Pale Ale
Magenta = Amber Ale
Cyan = Lager + Pilsener
Yellow = Belgians (light + dark)
Black = Stout + Porter
Blue = Everything else

Jamieson and Nowak '11

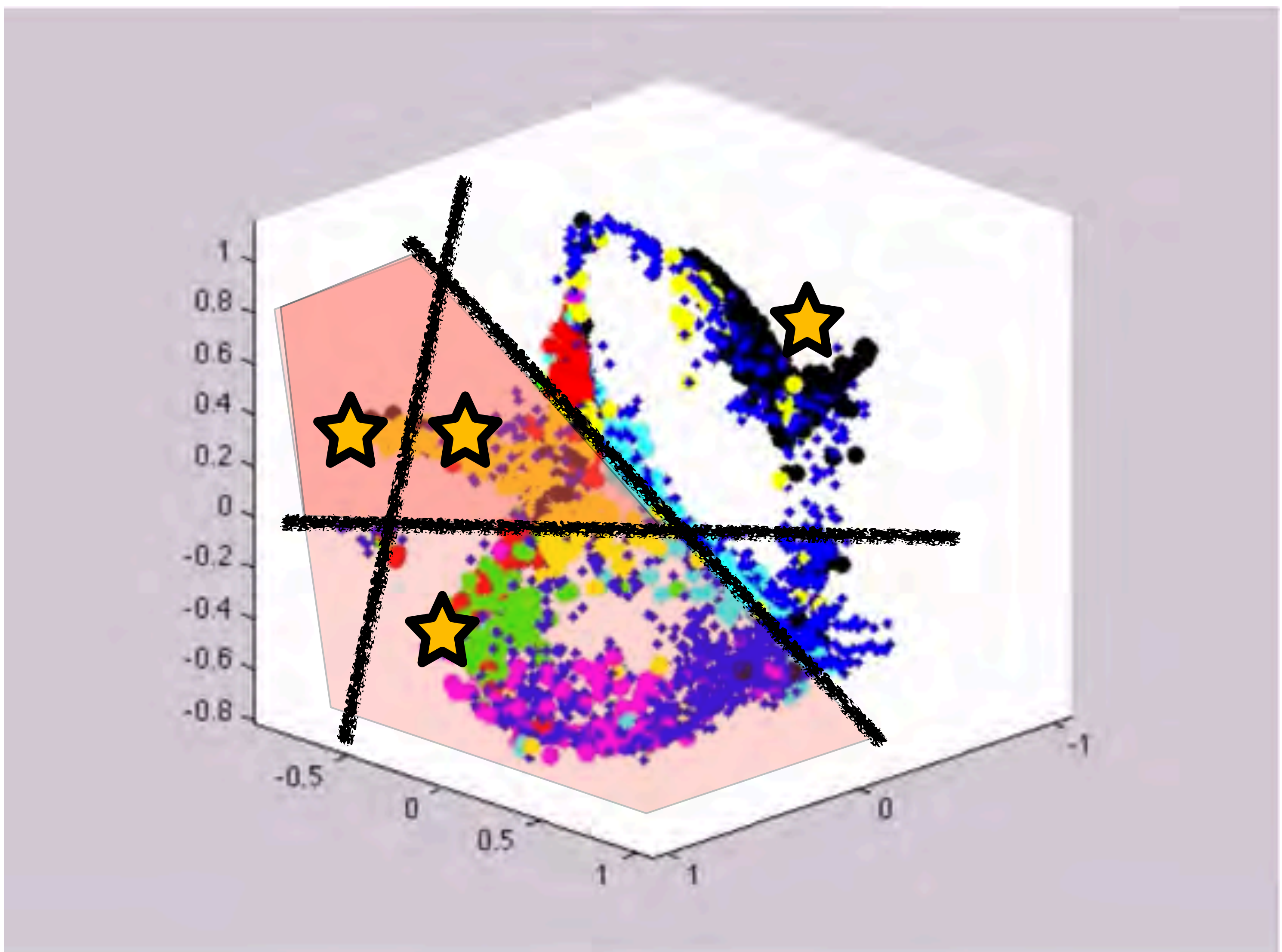
Ask AI to compare or rate strategically selected beers



Red = IPA
Green = Pale Ale
Magenta = Amber Ale
Cyan = Lager + Pilsener
Yellow = Belgians (light + dark)
Black = Stout + Porter
Blue = Everything else

Jamieson and Nowak ’11

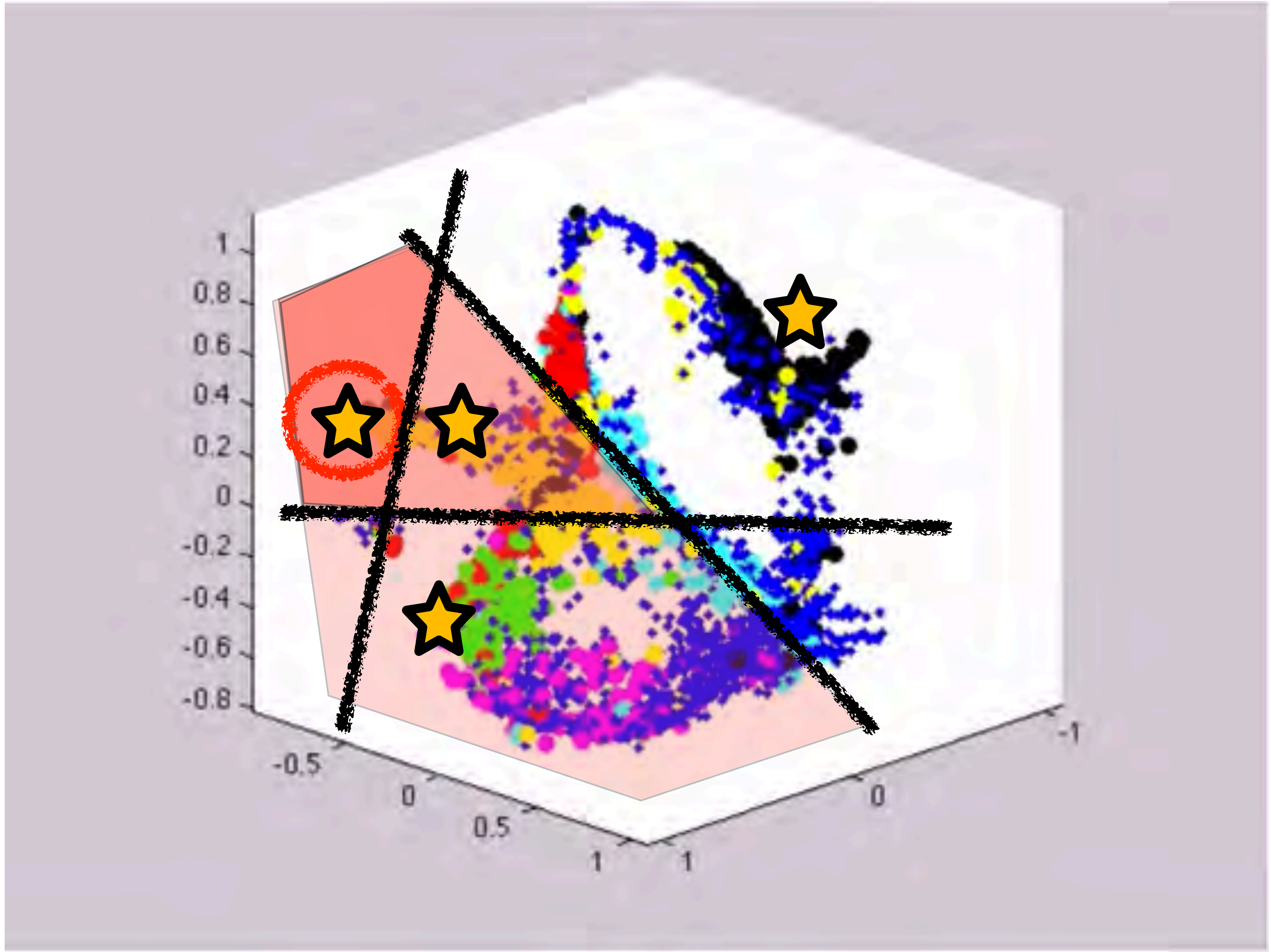
Ask AI to compare or rate strategically selected beers



Red = IPA
Green = Pale Ale
Magenta = Amber Ale
Cyan = Lager + Pilsener
Yellow = Belgians (light + dark)
Black = Stout + Porter
Blue = Everything else

Jamieson and Nowak '11

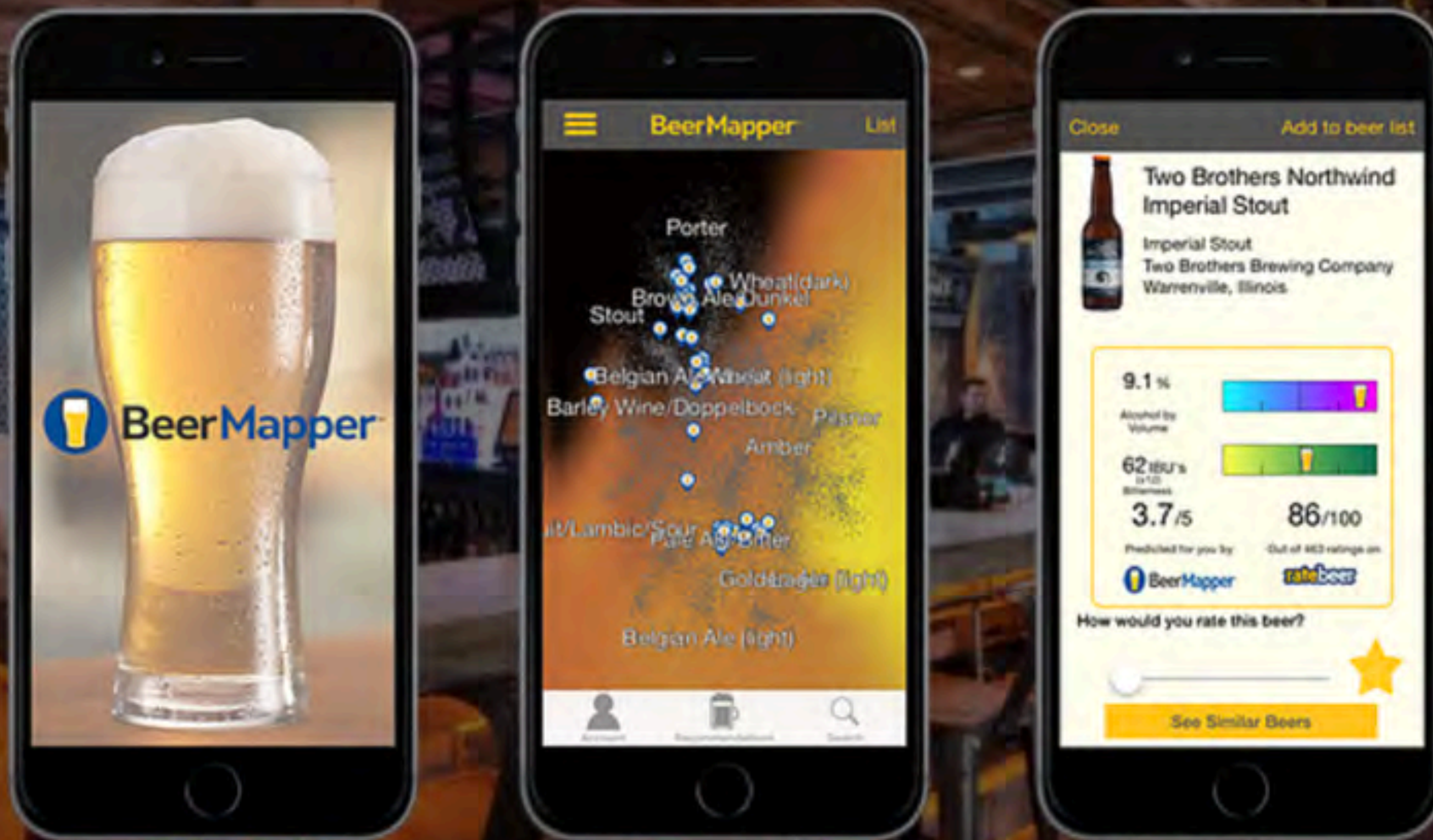
Ask AI to compare or rate strategically selected beers



Red = IPA
Green = Pale Ale
Magenta = Amber Ale
Cyan = Lager + Pilsener
Yellow = Belgians (light + dark)
Black = Stout + Porter
Blue = Everything else

Jamieson and Nowak '11

Discover better beer.



The most powerful beer app on the planet.


Cheers!