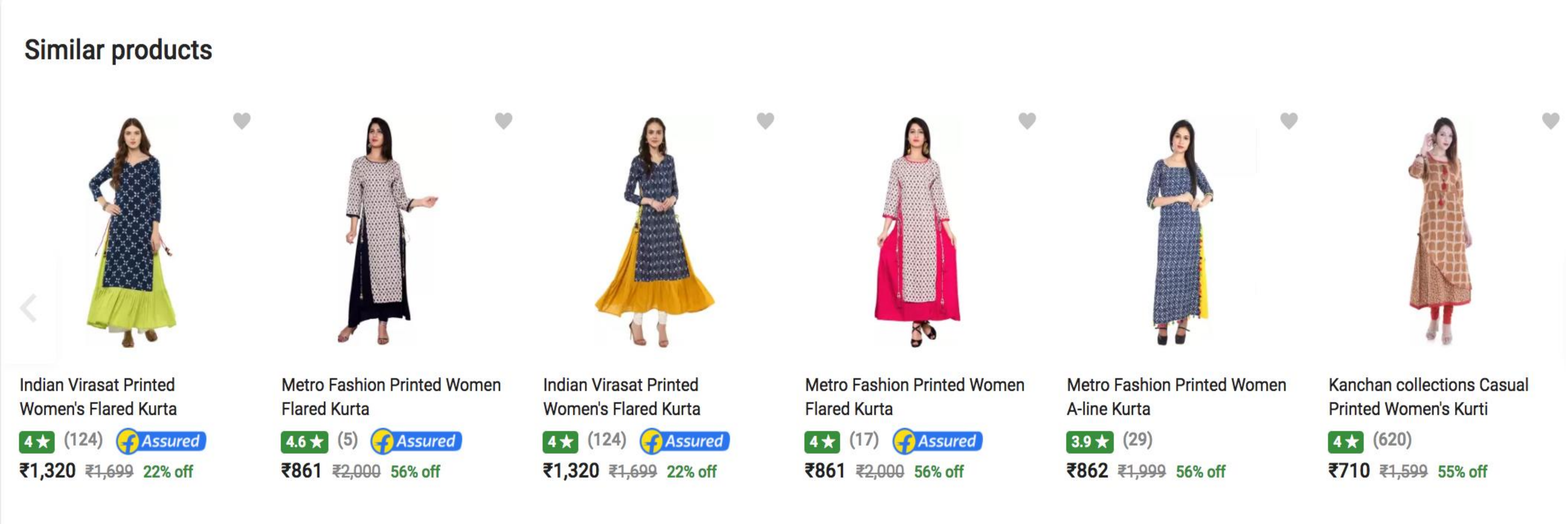
Distance-based Classification for Product Recommendation

Unit 6 Machine Learning

Product Recommendation

When we search a product, retail websites suggest similar products



Similar products

Indian Virasat Printed Women's Flared Kurta
4★ (124) Assured
₹1,320 ~~₹1,699~~ 22% off

Metro Fashion Printed Women Flared Kurta
4.6★ (5) Assured
₹861 ~~₹2,000~~ 56% off

Indian Virasat Printed Women's Flared Kurta
4★ (124) Assured
₹1,320 ~~₹1,699~~ 22% off

Metro Fashion Printed Women Flared Kurta
4★ (17) Assured
₹861 ~~₹2,000~~ 56% off

Metro Fashion Printed Women A-line Kurta
3.9★ (29)
₹862 ~~₹1,999~~ 56% off

Kanchan collections Casual Printed Women's Kurti
4★ (620)
₹710 ~~₹1,599~~ 55% off

Image source: Flipkart tech blog

Product Recommendation

When we search a product, retail websites suggest similar products

How does it know which products are “similar” to the original product?

What is even meant by “similarity of products”?

Product Recommendation

When we search a product, retail websites suggest similar products

How does it know which products are “similar” to the original product?

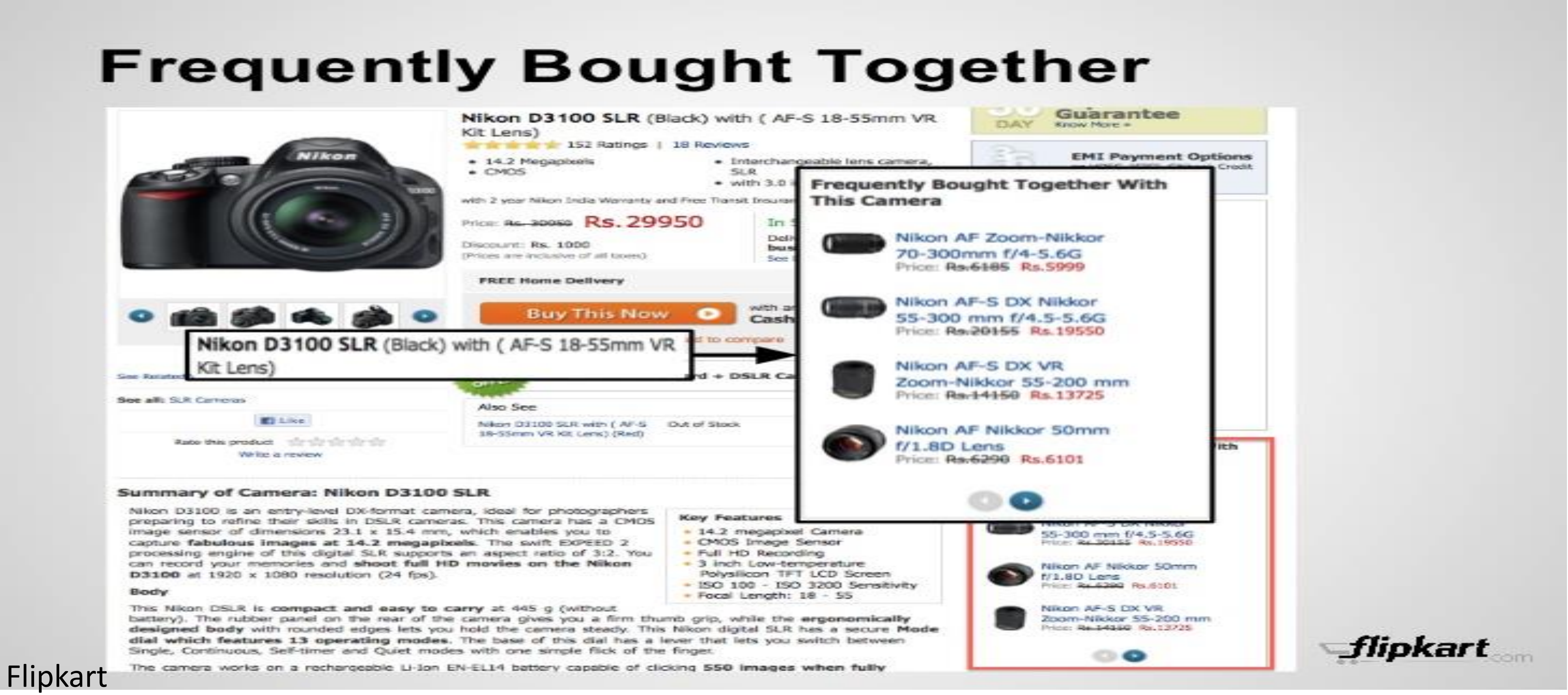
What is even meant by “similarity of products”?

Similarity notion 1: Based on customer statistics

Similarity notion 2: Based on features of products

Recommendation based on statistics

Two products which are often bought together may be considered “similar”



Frequently Bought Together

Nikon D3100 SLR (Black) with (AF-S 18-55mm VR Kit Lens)

★★★★★ 152 Ratings | 18 Reviews

- 14.2 Megapixels
- CMOS
- Interchangeable lens camera, SLR
- with 3.0

with 2 year Nikon India Warranty and Free Transit Insurance

Price: ~~Rs. 30950~~ Rs. 29950

Discount: Rs. 1000

(Prices are inclusive of all taxes)

FREE Home Delivery

Buy This Now

with an Cash

Add to compare

Nikon D3100 SLR (Black) with (AF-S 18-55mm VR Kit Lens)

See all: SLR Cameras

Like

Rate this product ☆☆☆☆☆

Write a review

Also See

Nikon D3100 SLR with (AF-S 18-55mm VR Kit Lens) (Red) Out of Stock

30 DAY Guarantee Know More »

EMI Payment Options

Frequently Bought Together With This Camera

Nikon AF Zoom-Nikkor 70-300mm f/4-5.6G
Price: ~~Rs.6185~~ Rs.5999

Nikon AF-S DX Nikkor 55-300 mm f/4.5-5.6G
Price: ~~Rs.20155~~ Rs.19550

Nikon AF-S DX VR Zoom-Nikkor 55-200 mm
Price: ~~Rs.14150~~ Rs.13725

Nikon AF Nikkor 50mm f/1.8D Lens
Price: ~~Rs.6290~~ Rs.6101

55-300 mm f/4.5-5.6G
Price: ~~Rs.20155~~ Rs.19550

Nikon AF Nikkor 50mm f/1.8D Lens
Price: ~~Rs.6290~~ Rs.6101

Nikon AF-S DX VR Zoom-Nikkor 55-200 mm
Price: ~~Rs.14150~~ Rs.13725

Summary of Camera: Nikon D3100 SLR

Nikon D3100 is an entry-level DX-format camera, ideal for photographers preparing to refine their skills in DSLR cameras. This camera has a CMOS image sensor of dimensions 23.1 x 15.4 mm, which enables you to capture **fabulous images at 14.2 megapixels**. The swift EXPEED 2 processing engine of this digital SLR supports an aspect ratio of 3:2. You can record your memories and **shoot full HD movies on the Nikon D3100** at 1920 x 1080 resolution (24 fps).

Key Features

- 14.2 megapixel Camera
- CMOS Image Sensor
- Full HD Recording
- 3 inch Low-temperature Polysilicon TFT LCD Screen
- ISO 100 - ISO 3200 Sensitivity
- Focal Length: 18 - 55

Body

This Nikon DSLR is **compact and easy to carry** at 445 g (without battery). The rubber panel on the rear of the camera gives you a firm thumb grip, while the **ergonomically designed body** with rounded edges lets you hold the camera steady. This Nikon digital SLR has a secure **Mode dial which features 13 operating modes**. The base of this dial has a lever that lets you switch between Single, Continuous, Self-timer and Quiet modes with one simple flick of the finger.

The camera works on a rechargeable Li-Ion EN-EL14 battery capable of clicking **550 images when fully**

Image source: Flipkart

Recommendation based on features

- Shirt and tie often bought together, but are they similar?
- What if the user is looking for choices among shirts?
- Product Type Matching The recommender system should show other shirts!

Recommended shirts should be “similar” to the shirt being viewed

Similarity defined in terms of attributes

Colour, size, brand, perhaps design and texture ………

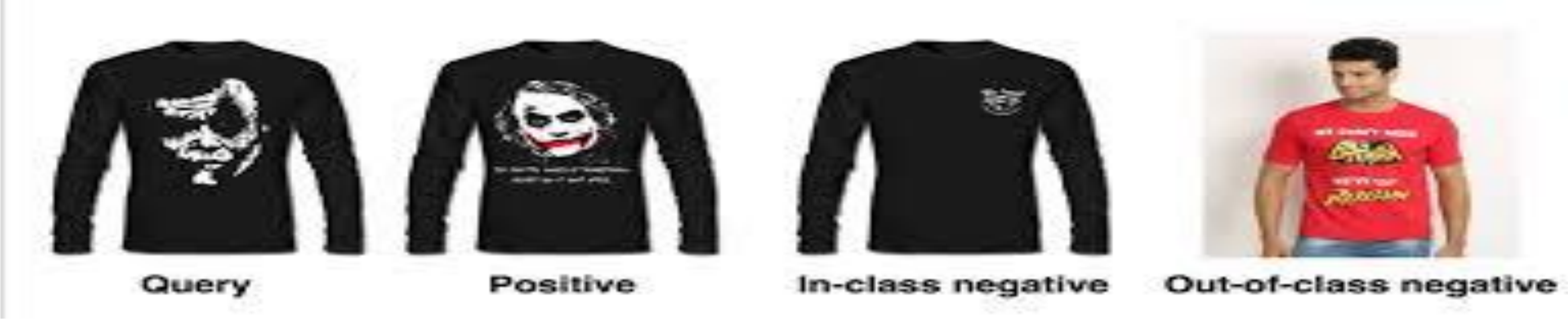


Image source: Google Images

Data Representation

- Represent each product by a vector of attributes or “features”
- Training data: (X_i, Y_i) where X: feature vector, Y: label
- Test data: ($X_{\text{test,}}$?) [Label to be predicted]

Data Representation

- Represent each product by a vector of attributes or “features”
- Training data: (X_i, Y_i) where X: feature vector, Y: label
- Test data: ($X_{test,}$?) [Label to be predicted]

- Label: can be binary, discrete or continuous
- Binary label: product recommendation (similar or not?)
- Discrete label: rating by one user (1star, 2star, 3star, 4star or 5star)?
- Continuous label: mean rating by many users (1,5)

A simple idea for classification

- Training data: (X_i, Y_i) where X: feature vector, Y: label
- Test data: (X_{test}, ?)
- X: representation of data-points (feature)

- Idea: things that “look” similar, are usually the same type!
- If X_{test} = X_i, then probably Y_{test} = Y_i !

But is it a good idea?

- Training data: (X_i, Y_i) where X: feature vector, Y: label
- Test data: (X_{test}, ?)
- X: representation of data-points (feature)

- Idea: things that “look” similar, are usually the same type!
- If X_{test} = X_i, then probably Y_{test} = Y_i !

- Is X a good enough representation?
- In reality, X_{test} = X_i will rarely happen (especially if X is continuous!)

But is it a good idea?

- Training data: (X_i, Y_i) where X: feature vector, Y: label
- Test data: (X_{test}, ?)
- X: representation of data-points (feature)

- Idea: things that “look” similar, are usually the same type!
- If X_{test} = X_i, then probably Y_{test} = Y_i !

- Is X a good enough representation? - Let’s assume it is
- In reality, X_{test} = X_i will rarely happen (especially if X is continuous!)

Distance between feature vectors

- If X is continuous-valued, X_$_{test}$ = X_i will almost surely never happen!
- But, X_$_{test}$ ~ X_i is possible!

- X_$_{test}$ ~ X_i : Euclidean Distance between the two points is very less
- || X_$_{test}$ - X_i ||$_2$ is very low!

- ||a – b|| $_2$ = √∑(a_i – b_i)2 (also called the l$_2$-norm of a-b)

Nearest-Neighbor Classification

- Training: N labelled examples (X_i, Y_i) where i: 1 to N
- Function learnt: the training set itself!
- Testing: X_test
- $Y\_{\text{pred}} = Y\_{n}$, where $n = \arg\min_{i} ||X\_{\text{test}} - X\_i||_{2}$

Nearest-Neighbor Classification

- Training: N labelled examples (X_i, Y_i) where i: 1 to N
- Function learnt: the training set itself!
- Testing: X_test
- $Y\_{pred} = Y\_{n}$, where $n = \arg\min_i ||X\_{test} – X\_i||_2$

- Compute the Euclidean distance between the test datapoint and each of the N training datapoints
- Choose that training point for which this distance is minimum (Nearest-Neighbor)
- Use its label as the predicted label of the test point!

Nearest-Neighbor Classification

- Training: N labelled examples (X_i, Y_i) where i: 1 to N
- Function learnt: the training set itself!
- Testing: X_test
- $Y\_{\text{pred}} = Y\_{n}$, where $n = \arg\min_i \|X\_{\text{test}} – X\_i\|_2$

- Not very robust due to outliers!
- Too much computation and storage required!

K-nearest Neighbors Classification

- Problem 1: The nearest neighbour of the test point may be outlier!

- Outliers are rare
- K nearest neighbors: unlikely to contain many outliers

- Solution:

1) Sort the training points according to distance from test point
2) Choose the first K training points
3) Predicted label = most frequent label among them!

K-nearest Neighbors Regression

- Problem 1: The nearest neighbour of the test point may be outlier!

- Outliers are rare
- K nearest neighbors: unlikely to contain many outliers

- Solution:

1) Sort the training points according to distance from test point
2) Choose the first K training points
3) Predicted label = mean of their labels!

Nearest Mean Classification

- Problem 2: Too much computation (N) and storage (N*(D+1)) required!

- One solution: keep only one representative from each class
- How to choose the representative?
- Mean of feature vectors all data-points in that class!

- Mean for class k: $\mu\_{}_{k} = \sum 1(Y\_{}_{i}=k) * X\_{}_{i}\ 1(Y\_{}_{i}=k) / \sum 1(Y\_{}_{i}=k)$
- Compare test point $X\_{}_{test}$ with each $\mu\_{}_{k}$ and choose label of the closest!
- $Y\_{}_{pred} = \text{argmin}_{k}\ ||X\_{}_{test} - \mu\_{}_{k}\ ||_{2}$

Nearest Mean Classification

- Problem 2: Too much computation (N) and storage (N*(D+1)) required!

- One solution: keep only one representative from each class
- How to choose the representative?
- Mean of feature vectors all data-points in that class!

- Mean for class k: $\mu_k = \sum 1(Y_i=k) * X_i \ 1(Y_i=k) / \sum 1(Y_i=k)$
- Compare test point X_{test} with each μ_k and choose label of the closest!
- $Y_{pred} = \text{argmin}_k \, ||X_{test} - \mu_k||_2$ [K comparisons instead of N]