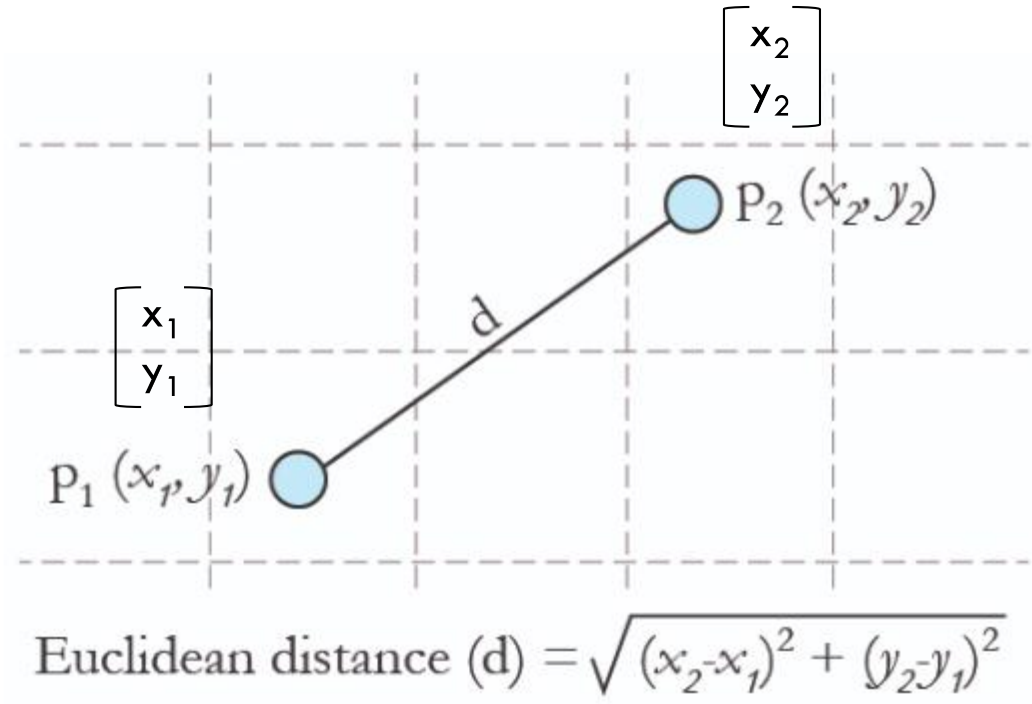
PROXIMITY MEASURES

Dr. Umarani Jayaraman
Assistant Professor

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY,
DESIGN AND MANUFACTURING,
KANCHEEPURAM

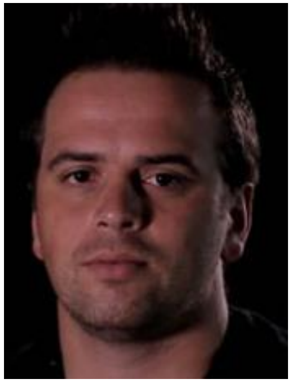
Introduction

- Euclidean Distance
  - To compute the distance between two points



Why do we need distance measure?

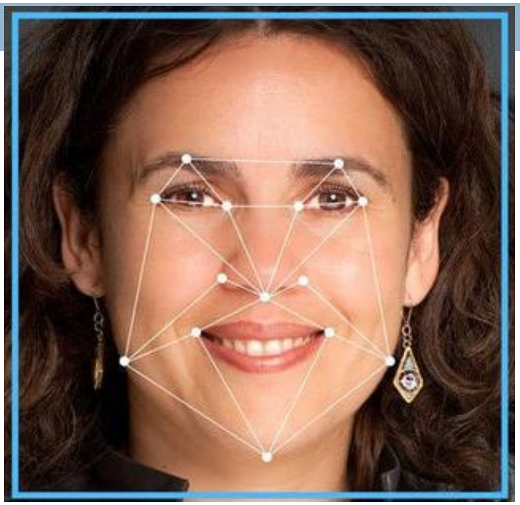
- Pattern Recognition
  - Recognizing similar patterns
  - Pattern is representation of an object with features/attributes
  - It is an important aspect of visual perception
- How does pattern recognition work?




Training samples

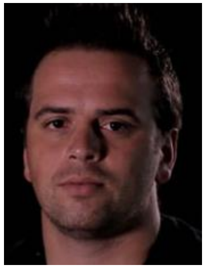
Testing samples

Motivation for distance measure



- How does pattern recognition work?
  - Pattern is represented using features
  - Feature vector: collection of features
- Features
  - Distance between the eyes
  - Width of the nose
  - Length of jaw line
  - Structure of the cheek

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}$$



$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix}$$

Compute the distance between these vectors to find similarity

Motivation for distance measure

- Why do we need distance measures?
  - To compare patterns during the task of pattern recognition
  - Many other applications

d-dimensional feature representation

- Feature vectors of dimensionality *d* representing two patterns

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_d \end{bmatrix} \qquad Y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_d \end{bmatrix}$$

- Representation of features of multiple objects
  - Consider database of N patterns/objects
  - Each pattern represented using feature vector of dimensionality d
  - So all the patterns can be collectively represented as a matrix D of size dx N

d-dimensional feature representation

Database contains N objects

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{1N} \\ x_{21} & x_{22} & x_{23} & x_{2N} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ x_{d1} & x_{d2} & x_{d3} & x_{dN} \end{bmatrix}_{dxN}$$

Test Sample

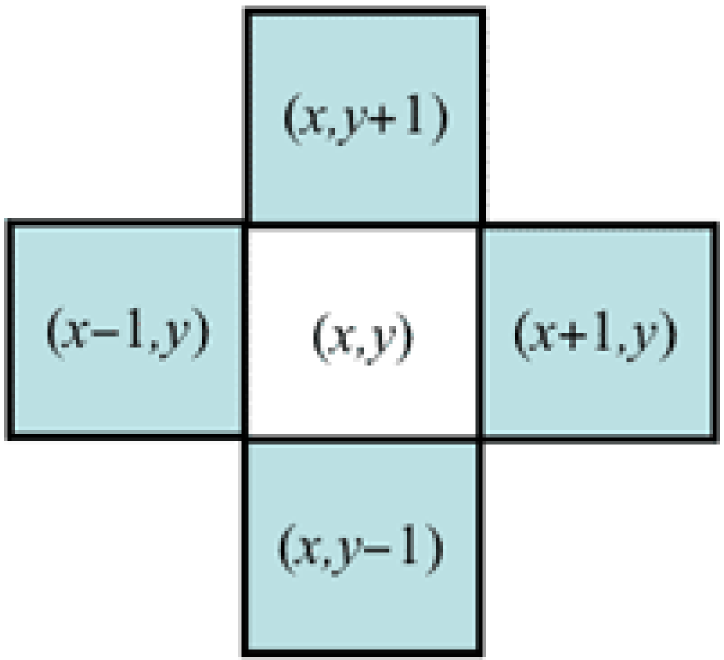
$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ yd \end{bmatrix}_{dx1}$$

L_p Norm

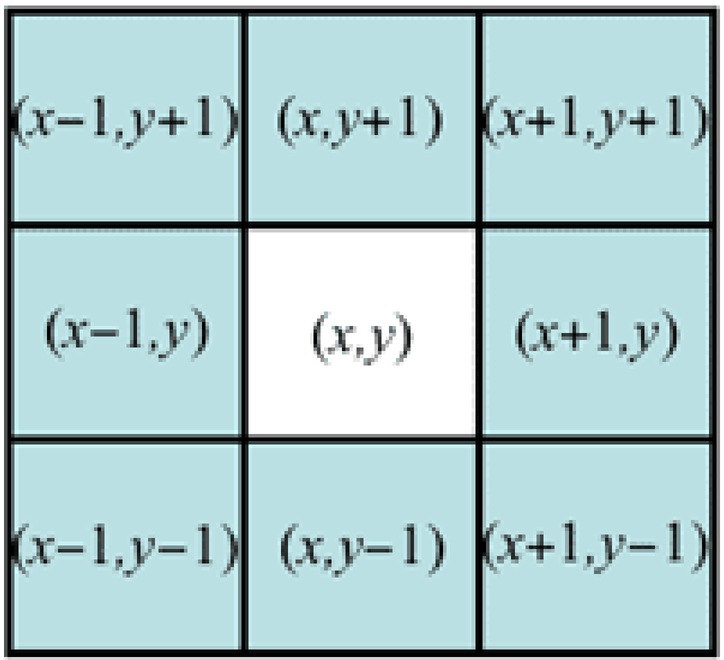
$$L_p(X,Y) = \left(\sum_{i=1}^{d} (| x_i - y_i |)^p \right)^{\frac{1}{p}}$$

- Where $p=1,2,\ldots,\infty$ and d is the dimension
- Depending on the value of p, we get different distance measures.
  - L_1 : Manhattan (city block distance)
  - L_2 : Euclidean
  - L_∞ : Max (chess board distance)
  - $L_{-\infty}$: Min
- This is also called as *Minkowski Norm*

Distance Measures



4-neighbourhood



8-neighbourhood

Distance Measures

		$(i, j+2)$ $(0,2)$		
	$(i-1, j+1)$ $(-1,1)$	$(i, j+1)$ $(0,1)$	$(i+1, j+1)$ $(1,1)$	
$(i-2, j)$ $(-2,0)$	$(i-1, j)$ $(-1,0)$	(i, j) $(0,0)$	$(i+1, j)$ $(1,0)$	$(i+2, j)$ $(2,0)$
	$(i-1, j-1)$ $(-1,-1)$	$(i, j-1)$ $(0,-1)$	$(i+1, j-1)$ $(1,-1)$	
		$(i, j-2)$ $(0,-2)$		

L_1 Norm/ Manhattan Distance

- When $p = 1$, in L_p norm we get the L_1 norm
- It is also called as
  - *Manhattan distance*
  - *City block distance.*
  - Taxi-cab distance
- This can be written as:

$$D(X, Y) = (\sum_{i=1}^{d} | x_i - y_i |)$$

L1 norm called as City Block Distance/ taxi-cab distance: Based on 4-connectivity

		2		
	2	1	2	
2	1	0	1	2
	2	1	2	
		2		

L_2 Norm/Euclidean Distance

- When $p = 2$, in L_p norm, we get L_2 norm
- It is also called as
  - Euclidean distance

$$D(X,Y) = \left(\sum_{i=1}^{d} | x_i - y_i |^2 \right)^{\frac{1}{2}}$$

L_∞ norm called as Chessboard Distance: Based on 8-connectivity

2	2	2	2	2
2	1	1	1	2
2	1	0	1	2
2	1	1	1	2
2	2	2	2	2

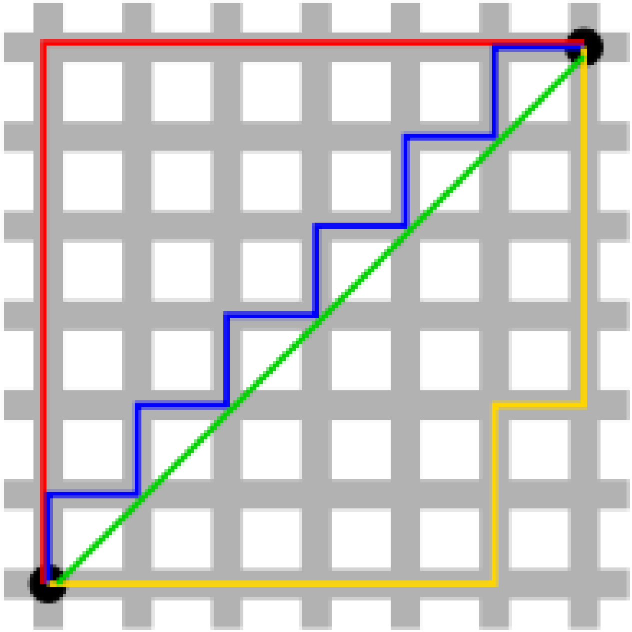
L_∞ Norm and $L_{-\infty}$ Norm

- The L_∞ norm is : It is also called as Chybyshev Distance
- $$D(X,Y) = max\left(\sum_{i=1}^{d} |x_i - y_i|\right)$$

- The $L_{-\infty}$ norm is :

$$D(X,Y) = min\left(\sum_{i=1}^{d} |x_i - y_i|\right)$$

Taxicab or city block distance vs Euclidean Distance



An illustration comparing the taxicab metric to the Euclidean metric on the plane: According to the taxicab metric the red, yellow, and blue paths have the same length which is 12. According to the Euclidean metric, the green path has length $6\sqrt{2} \approx 8.49$, and is the unique shortest path.

Distance Measures

- A distance measure is used to find the similarity between pattern representations.
- The distance function could be metric or non-metric.

Properties of a Metric Distance

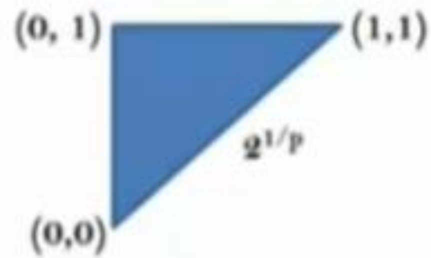
- A metric is a measure for which the following properties hold:
  - Non-negativity: D (X,Y) $\geq$ 0, $\forall X, Y$
  - Identity : $D(X,Y) = \ 0\, if(X = Y), \forall\, X, Y$
  - Symmetry : $D(X,Y) = D(Y,X) \forall\, X, Y$
  - Triangular Inequality: $D(X,Z) \leq D(X,Y)\ + D(Y,Z)\ \forall X, Y, Z$

  where $D(X,Y)$ gives the distance between X and Y.

Is L_p norm a metric?

- Is L_p norm a metric?
  - Yes, $\forall\, 1 \leq p \leq \infty$
  - No, if p <1

- When p<1, it does not satisfy the triangular inequality



  - Example:
    - the distance between (0,0) and (1,1) is $2^{1/p} > 2$, but the point (0,1) is at a distance 1 from both of these points.
    - Since this violates the triangle inequality for p<1, Minkowski distance is not a metric distance for p<1

Summary- Metric Similarity Function

- Motivation
- Minkowski Norm/ L_p Norm
- Properties of Metric
- Euclidean

Summary

- Euclidean Distance:

$$D_e(p, q) = [(x-s)^2 + (y-t)^2]^{1/2}$$

- City Block (Manhattan) Distance:

$$D_4(p, q) = |x-s| + |y-t|$$

- Chess Board (max) Distance:

$$D_8(p, q) = max(|x-s|, |y-t|)$$

		2		
	2	1	2	
2	1	0	1	2
	2	1	2	
		2		

City Block

2	2	2	2	2
2	1	1	1	2
2	1	0	1	2
2	1	1	1	2
2	2	2	2	2

Chessboard Block

THANK YOU

Source-Histogram distance

- https://stats.stackexchange.com/questions/7400/how-to-assess-the-similarity-of-two-histograms

- https://mpatacchiola.github.io/blog/2016/11/12/the-simplest-classifier-histogram-intersection.html