KNN - for Classification

Supervised

Hisham Ihshaish
Bristol, UK
June, 2021

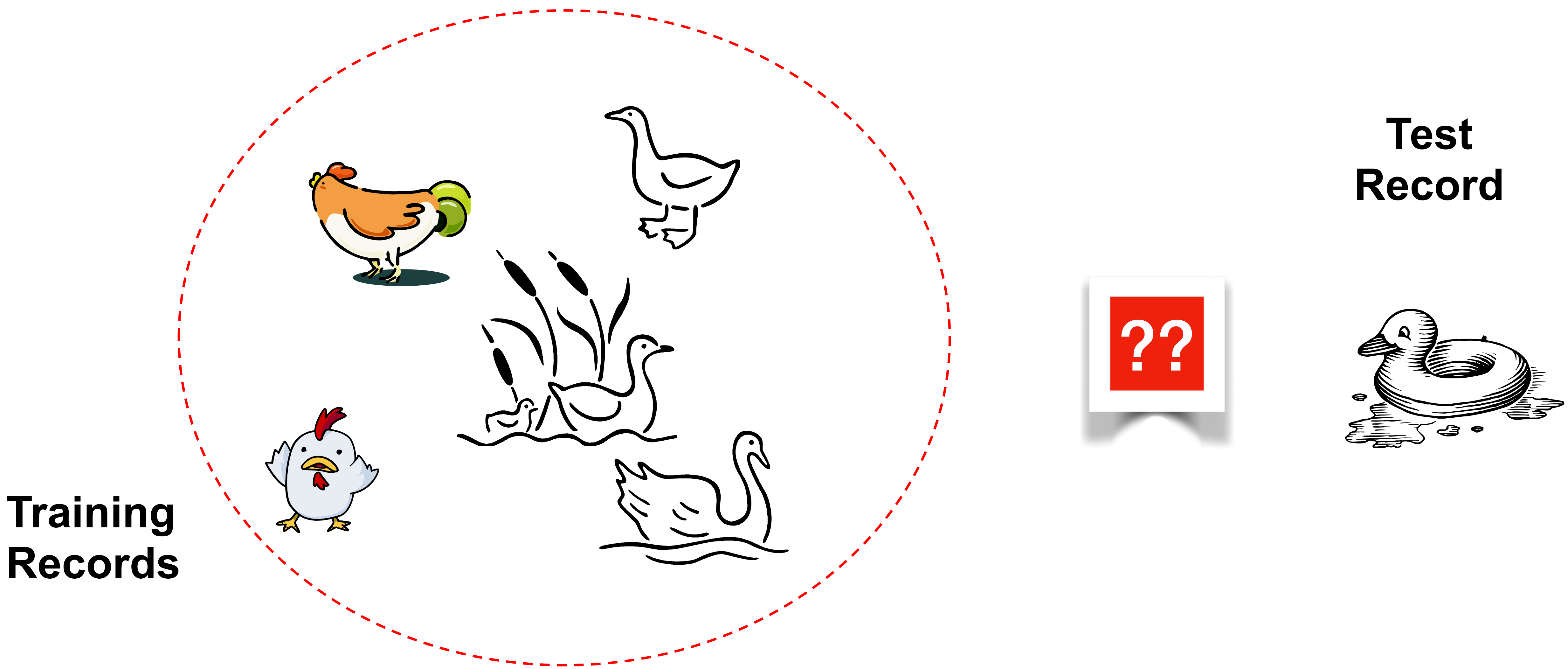
Agenda

➡ K Nearest Neighbour/**Neighbor - kNN**

- Properties
- Notes/thoughts
- Limitations

➡ Exercise - application of KNN for classification

Nearest Neighbour Classifiers



International Statistical Review (1989), **57**, 3, pp. 233–247. Printed in Great Britain

E. Fix and J.L. Hodges (1951): An Important Contribution to Nonparametric Discriminant Analysis and Density Estimation

Commentary on Fix and Hodges (1951)

B.W. Silverman and M.C. Jones

School of Mathematical Sciences, University of Bath, Bath, BA2 7AY, U.K.

Summary

In 1951, Evelyn Fix and J.L. Hodges, Jr. wrote a technical report which contained prophetic work on nonparametric discriminant analysis and probability density estimation, and which was never published by the authors. The report introduced several important concepts for the first time. It is of interest not only for historical reasons but also because it contains much material that is still

Read more about Evelyn here.

Hasan Ibn al-Haytham (Alhazen)

Born	c. 965 (c. 354 AH)[1] Basra, Iraq
Died	c. 1040 (c. 430 AH)[2] Cairo, Egypt
Residence	Basra · Cairo
Known for	*Book of Optics*, *Doubts Concerning Ptolemy*, Alhazen's problem, Analysis,[3] Catoptrics,[4] Horopter, Moon illusion, experimental science, scientific methodology,[5] visual perception, empirical theory of perception, Animal psychology[6]
Scientific career	
Fields	Optics · Astronomy · Mathematics

ELSEVIER

Pattern Recognition Letters

Volume 38, 1 March 2014, Pages 34-37

Alhazen and the nearest neighbor rule

Marcello Pelillo

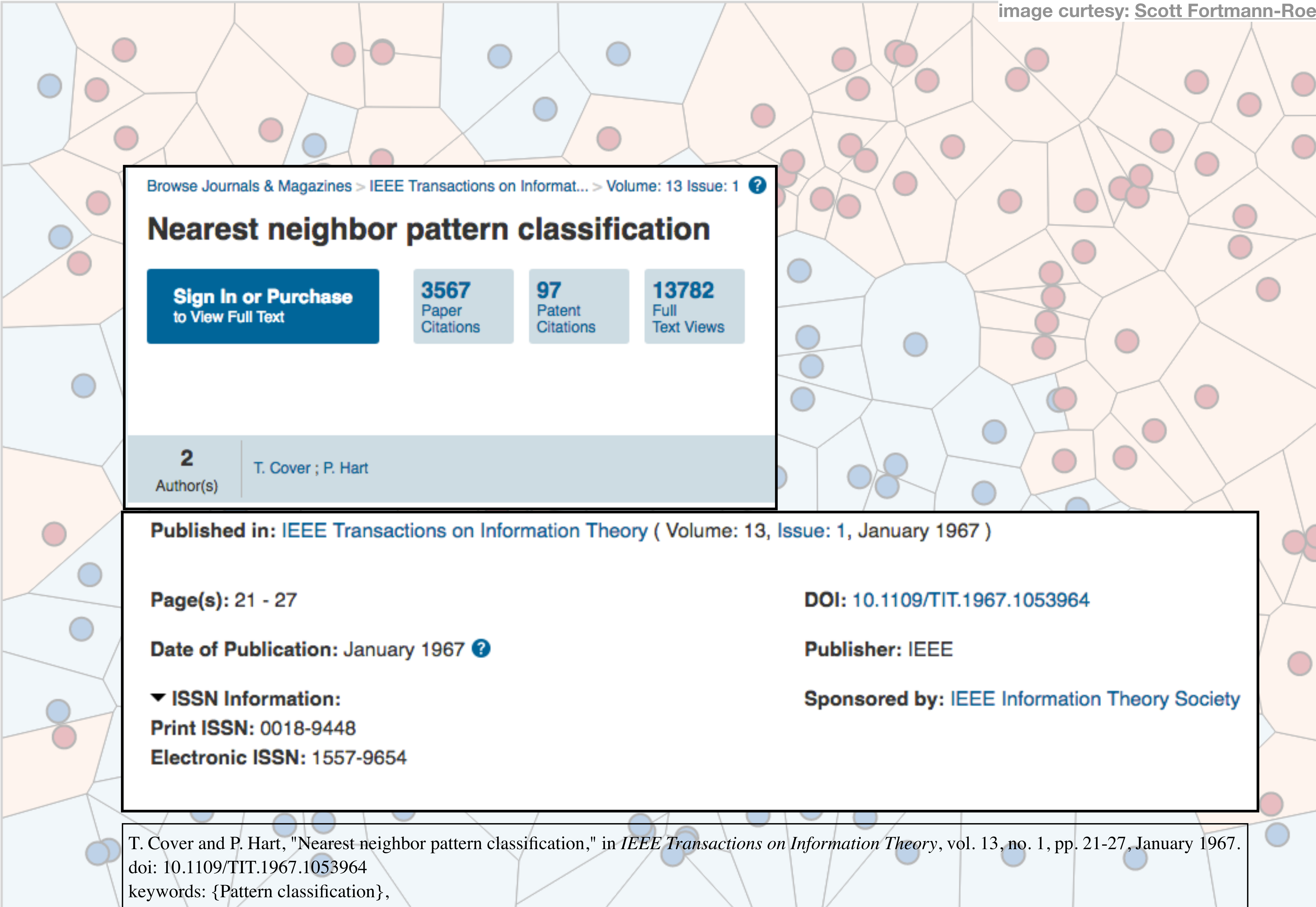
Show more

https://doi.org/10.1016/j.patrec.2013.10.022

Get rights and content

“Who invented the nearest neighbor rule?”, Pattern Recognition Tools

Marcello Pelillo, "Alhazen and the nearest neighbor rule" image curtesy: Scott Fortmann-Roe



Browse Journals & Magazines > IEEE Transactions on Informat... > Volume: 13 Issue: 1

Nearest neighbor pattern classification

Sign In or Purchase to View Full Text

2 Author(s) — T. Cover ; P. Hart

Published in: IEEE Transactions on Information Theory (Volume: 13, Issue: 1, January 1967)

Page(s): 21 - 27

Date of Publication: January 1967

▼ **ISSN Information:**
Print ISSN: 0018-9448
Electronic ISSN: 1557-9654

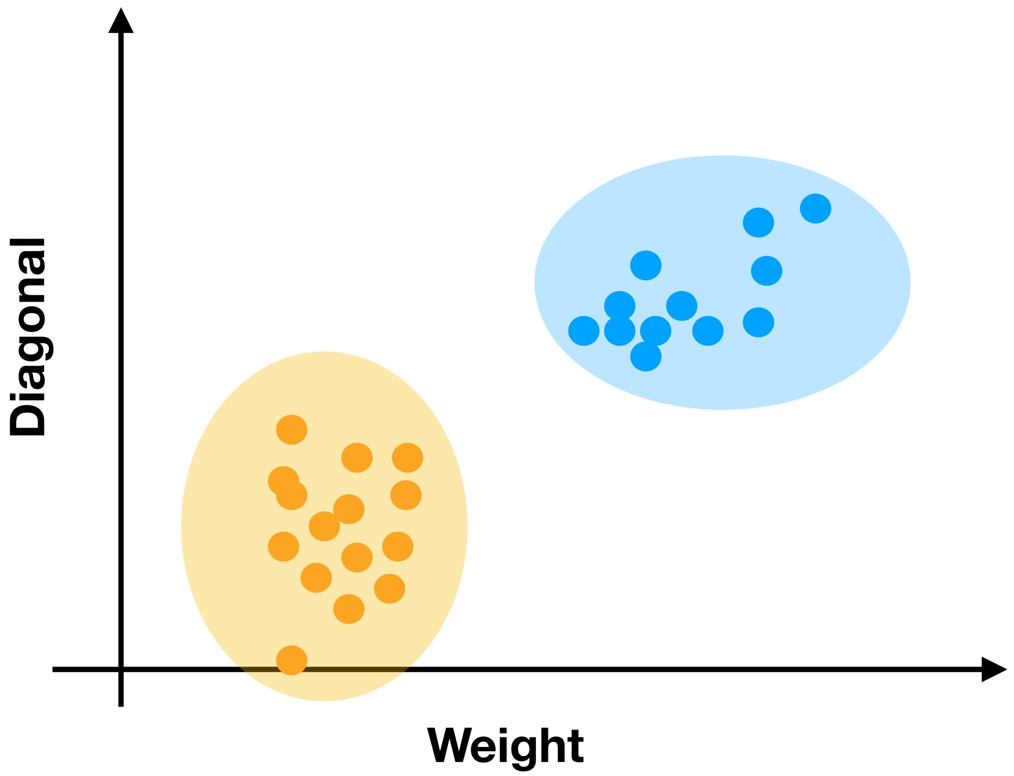
DOI: 10.1109/TIT.1967.1053964

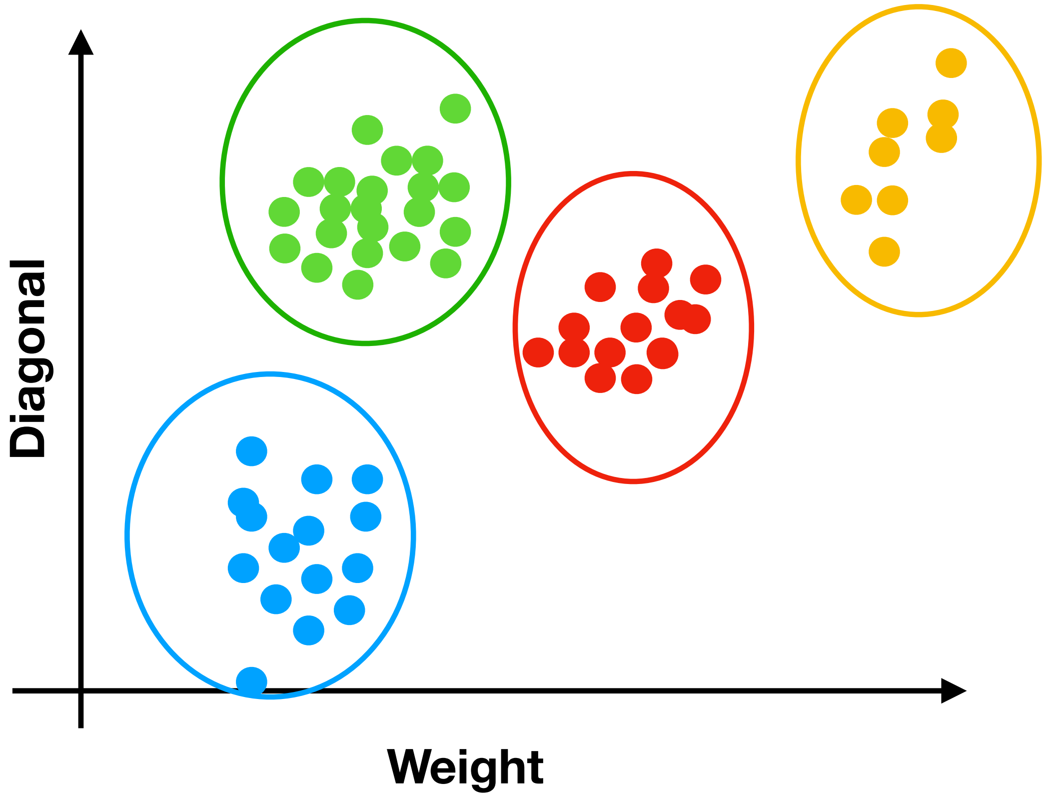
Publisher: IEEE

Sponsored by: IEEE Information Theory Society

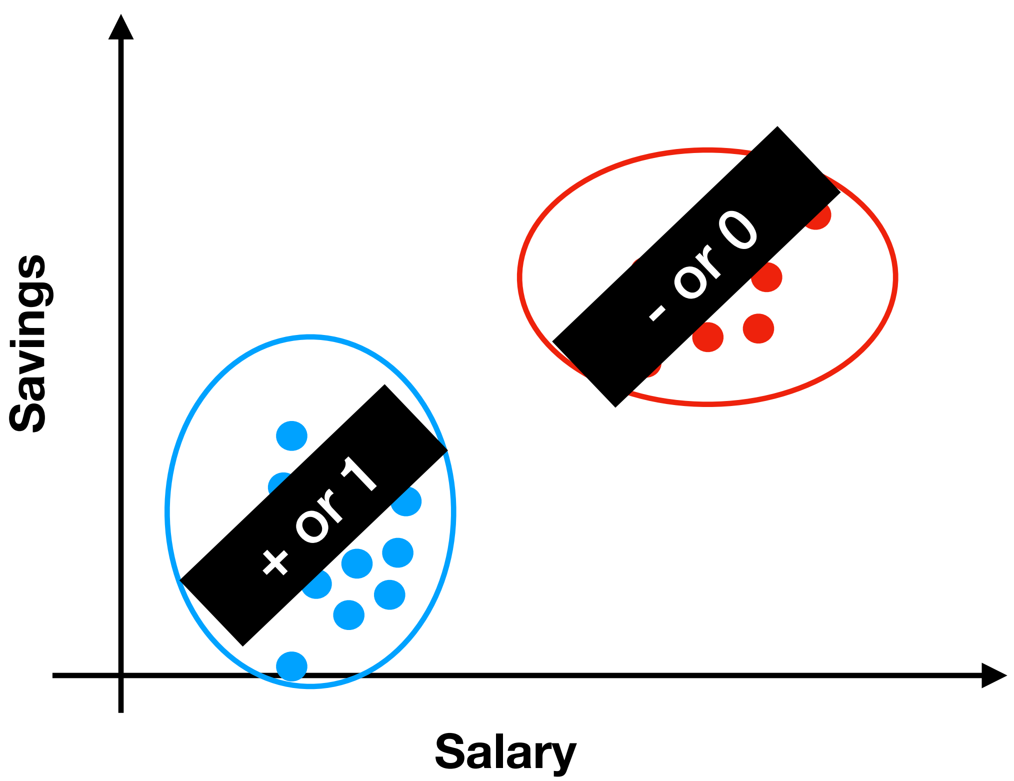
T. Cover and P. Hart, "Nearest neighbor pattern classification," in *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21-27, January 1967.
doi: 10.1109/TIT.1967.1053964
keywords: {Pattern classification},
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1053964&isnumber=22633

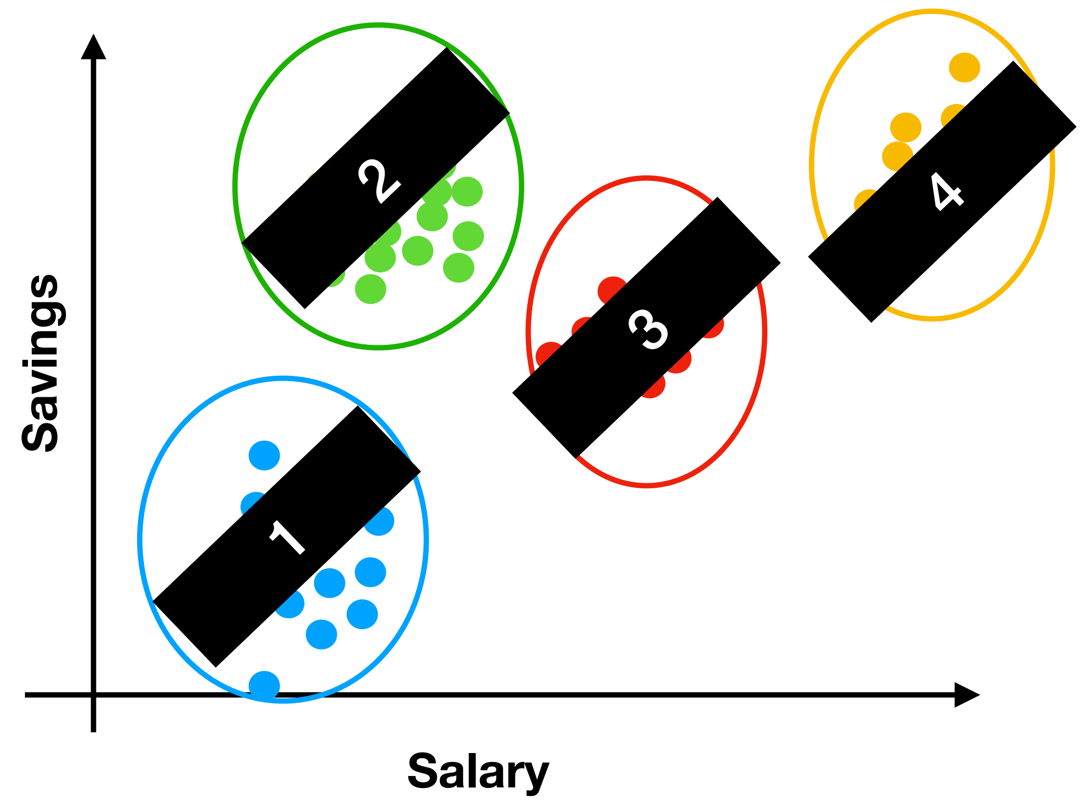
Binary vs Muti-class





Binary vs Muti-class

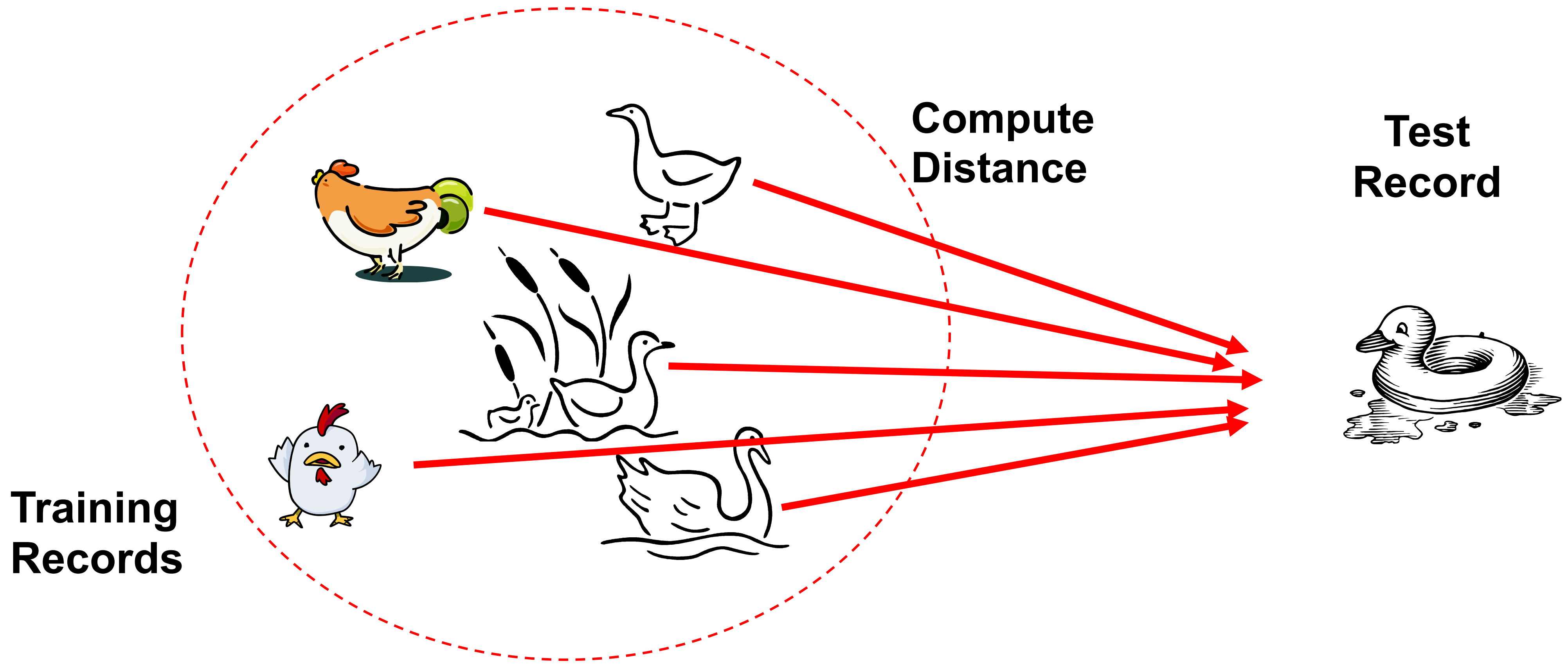




Nearest Neighbour Classifiers

Basic Idea

It walks like a duck, quacks like a duck, then it's probably a duck!



Agenda

- K Nearest Neighbour/**Neighbor - kNN**
  - **Properties**
  - Notes/thoughts
  - Limitations
- Exercise - application of KNN for classification

‘Lazy’ Learning

- It does not build a model explicitly - *instance-based*.
- No model ➡ no explicit hypothesis: *non parametric*!
  - ➡ many local (*implicit*) hypotheses - richer compared to the one hypothesis as in parametric learning.
- Classifying new instances is relatively expensive!
- Training is cheap, well, because there isn’t training!

*K*NN altogether

$$\boxed{y(x') = y(x) \text{ where } x = \underset{x \in X}{\operatorname{argmin}}\, dist(x, x')}$$

*K*NN altogether

$$y(x') = y(x) \text{ where } x = \underset{x \in X}{\operatorname{argmin}}\ dist(x, x')$$

class for new instance x' = class for known instance x

where x is the closest instance to x'

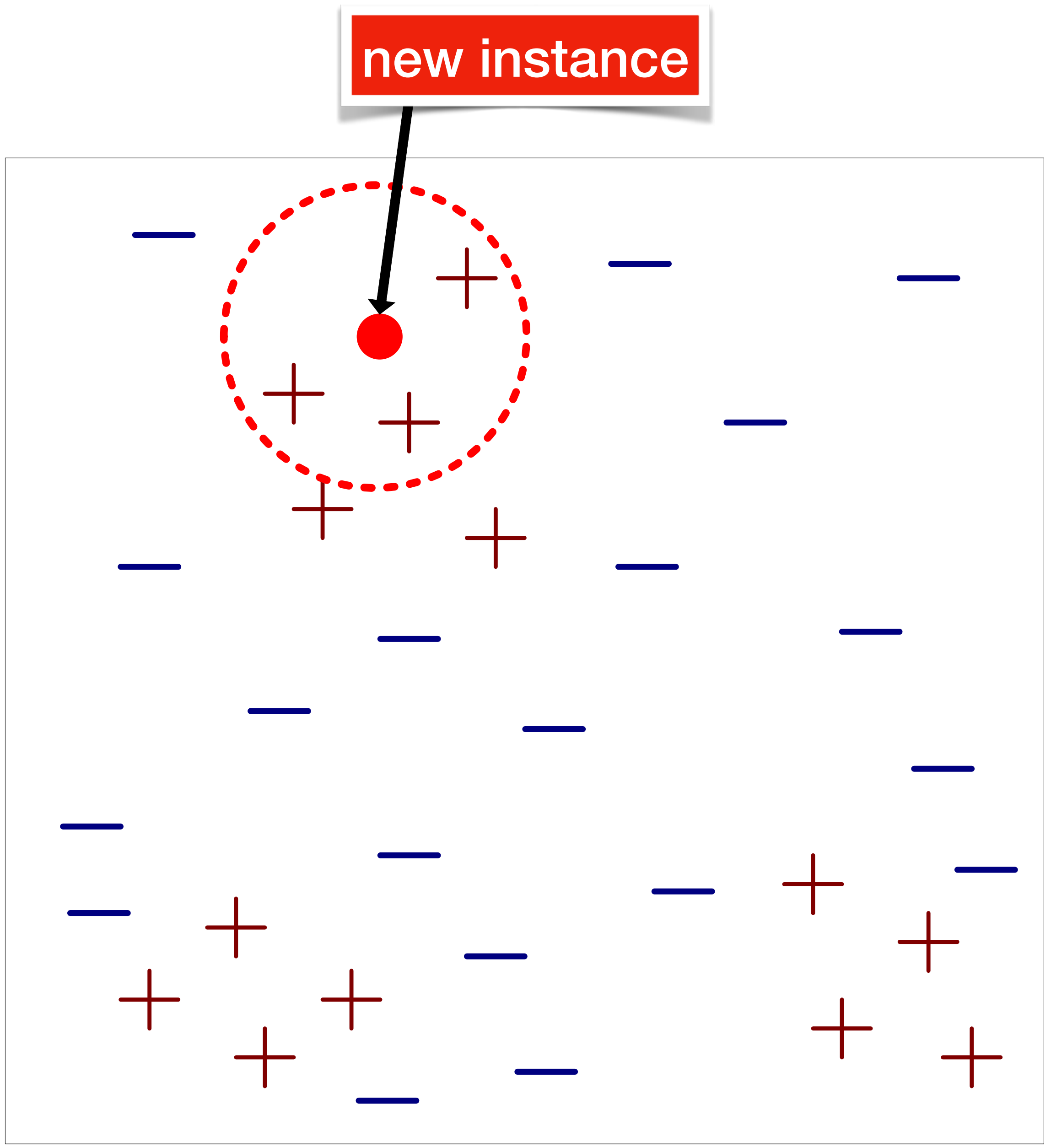
ie, the instance x with minimum distance to x'

find closest example in training set, and return corresponding label

*K*NN altogether

Requirements

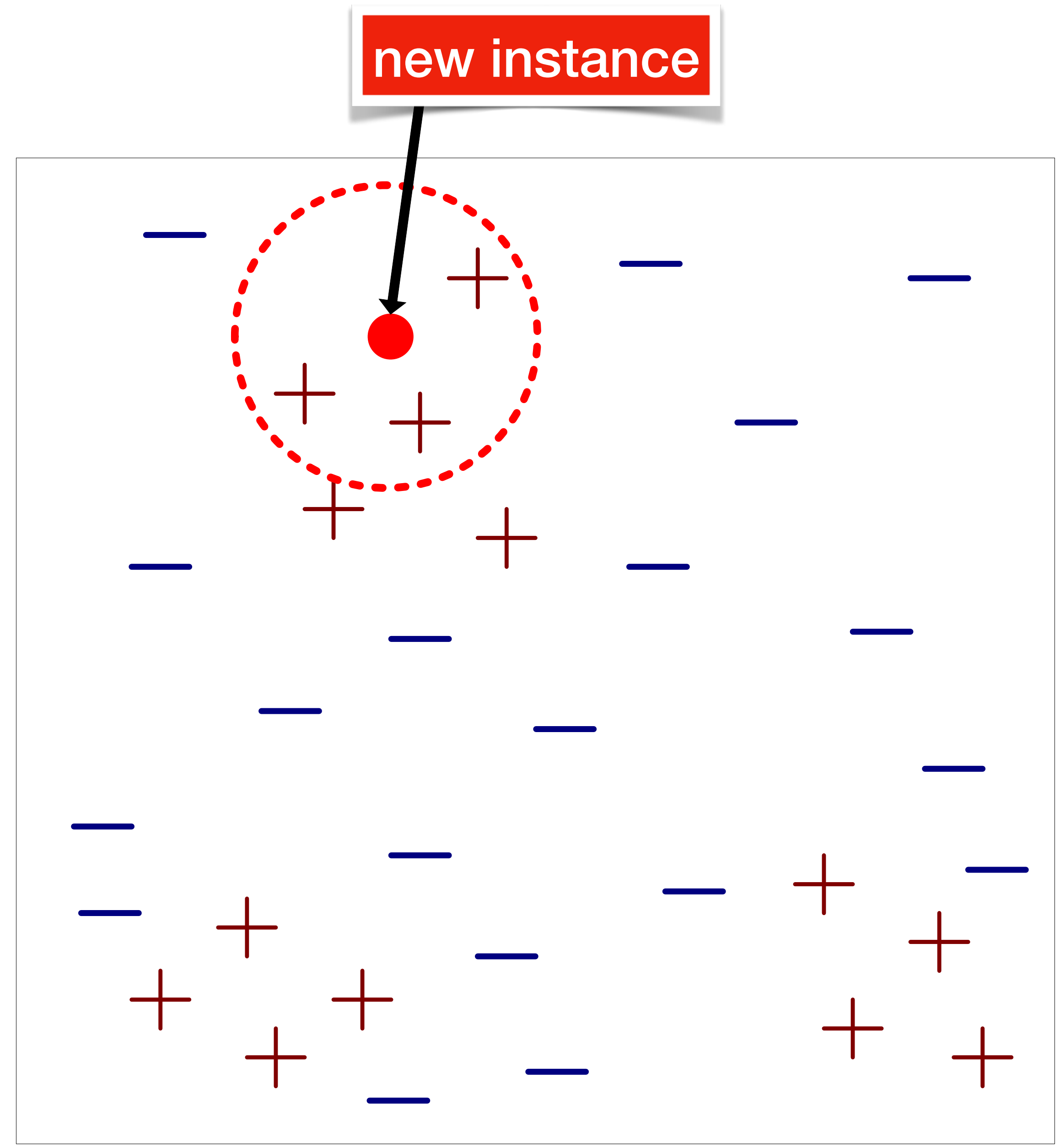
- Set of observations (stored samples and their labels).
- Distance Metric
- Value for *K*; the number of nearest neighbours to consider!



*K*NN altogether

Process

- Compute ***distance*** of new instance to all other instances.
- Identify *K* nearest neighbours.
- Use class labels of nearest neighbours to determine class label of the new (unknown) instance.



Detailed…

$$P(y=j|X=x)=\frac{1}{K}\sum_{i\in\mathcal{A}} I(y^{(i)}=j)$$

1

- It runs through the whole dataset computing ***d*** between ***x*** and each “training” instance. We’ll call the ***K*** points in the training data that are closest to ***x*** the set ***A***.

2

- It then estimates the conditional probability for each class, that is, the fraction of points in ***A*** with that given class label (***j*** is checked here). (Note ***I****(****x****)* is the indicator function which evaluates to **1** when the argument **x** is true and **0** otherwise)

Detailed …

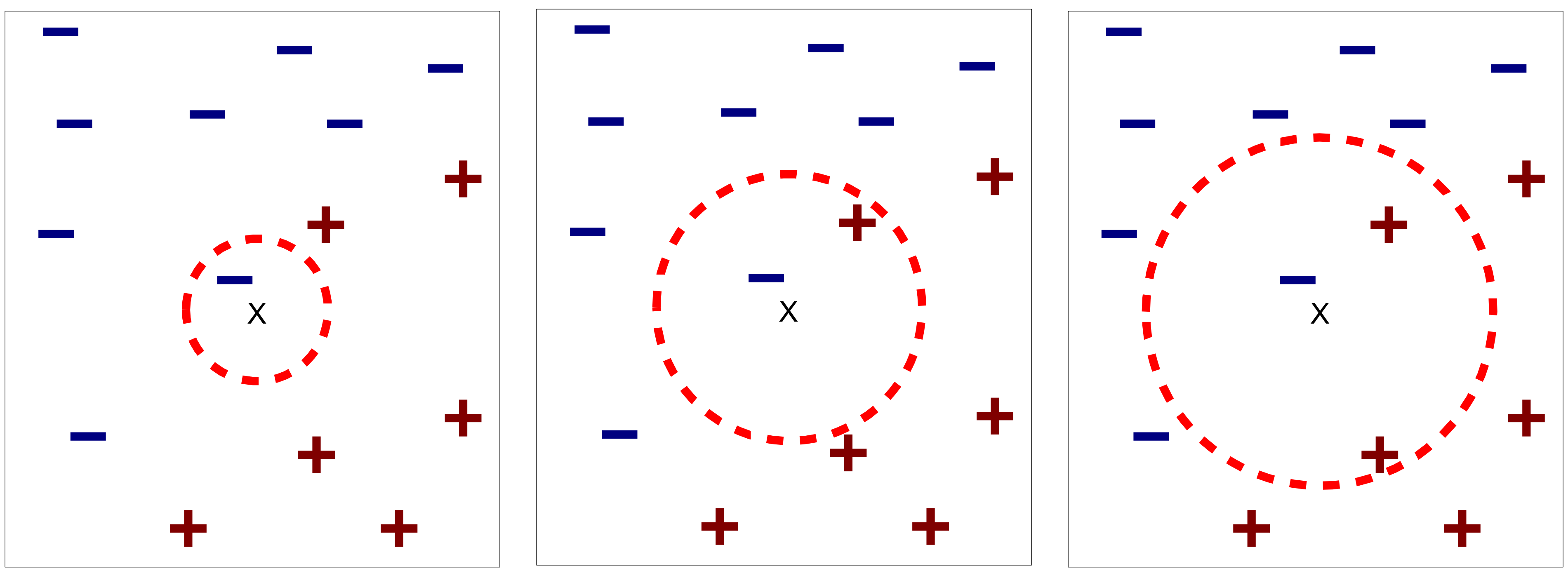
$$P(y = j | X = x) = \frac{1}{K} \sum_{i \in \mathcal{A}} I(y^{(i)} = j)$$

find closest examples (*K*) in training set, and return the label corresponding to *K* neighbours' majority vote!

The Unknowns …

- *K* (hyperparameter)
- Distance (similarity metric)

the K



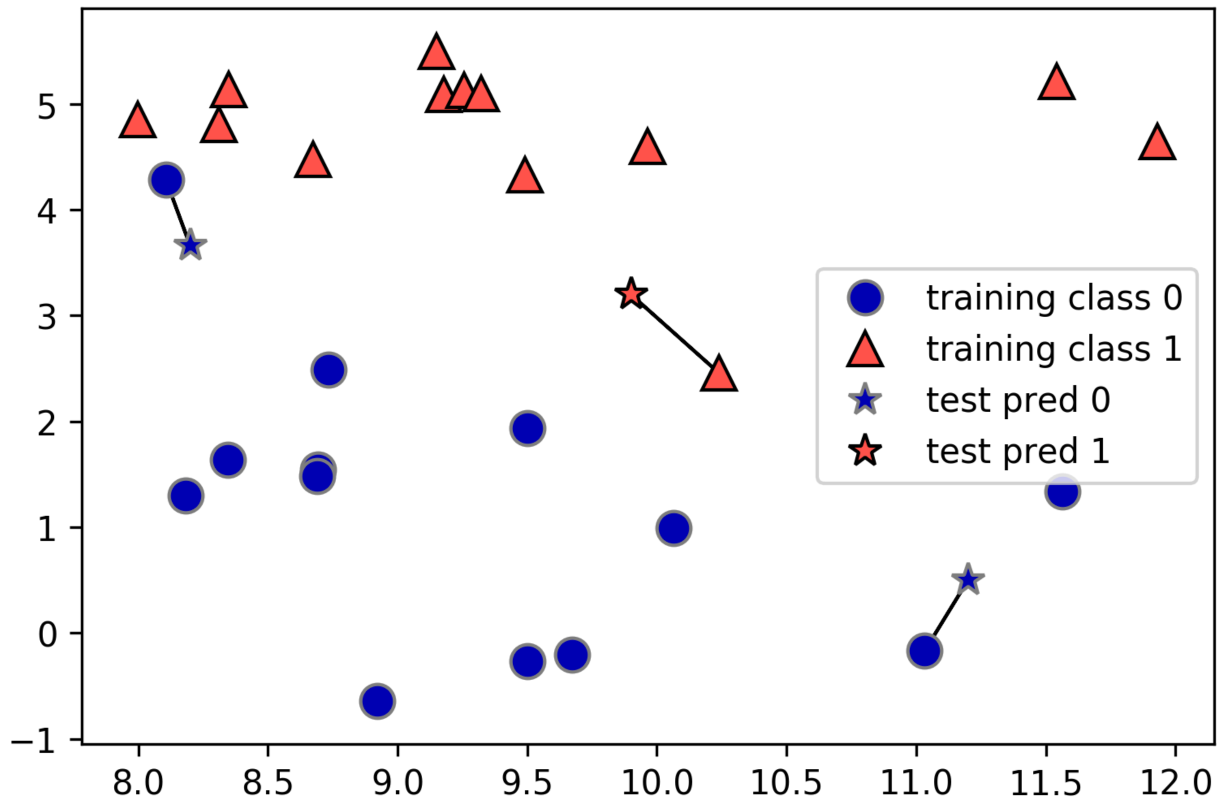
(a) 1-nearest neighbor **K =1**

(b) 2-nearest neighbor **K =2**

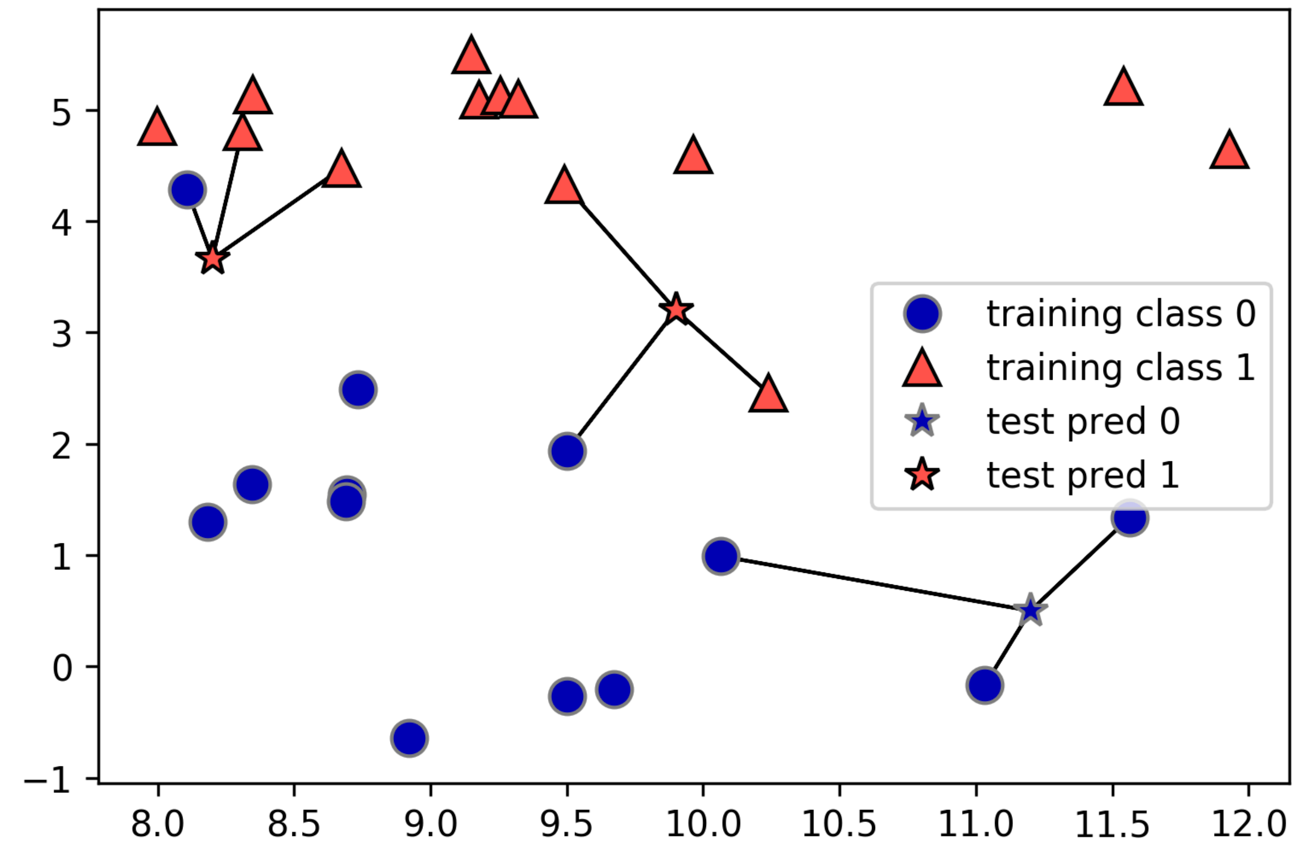
(c) 3-nearest neighbor **K =3**

k-nearest neighbours of an instance x are the k points (instances) closest to x.

the *K*



training class 0
training class 1
test pred 0
test pred 1



training class 0
training class 1
test pred 0
test pred 1

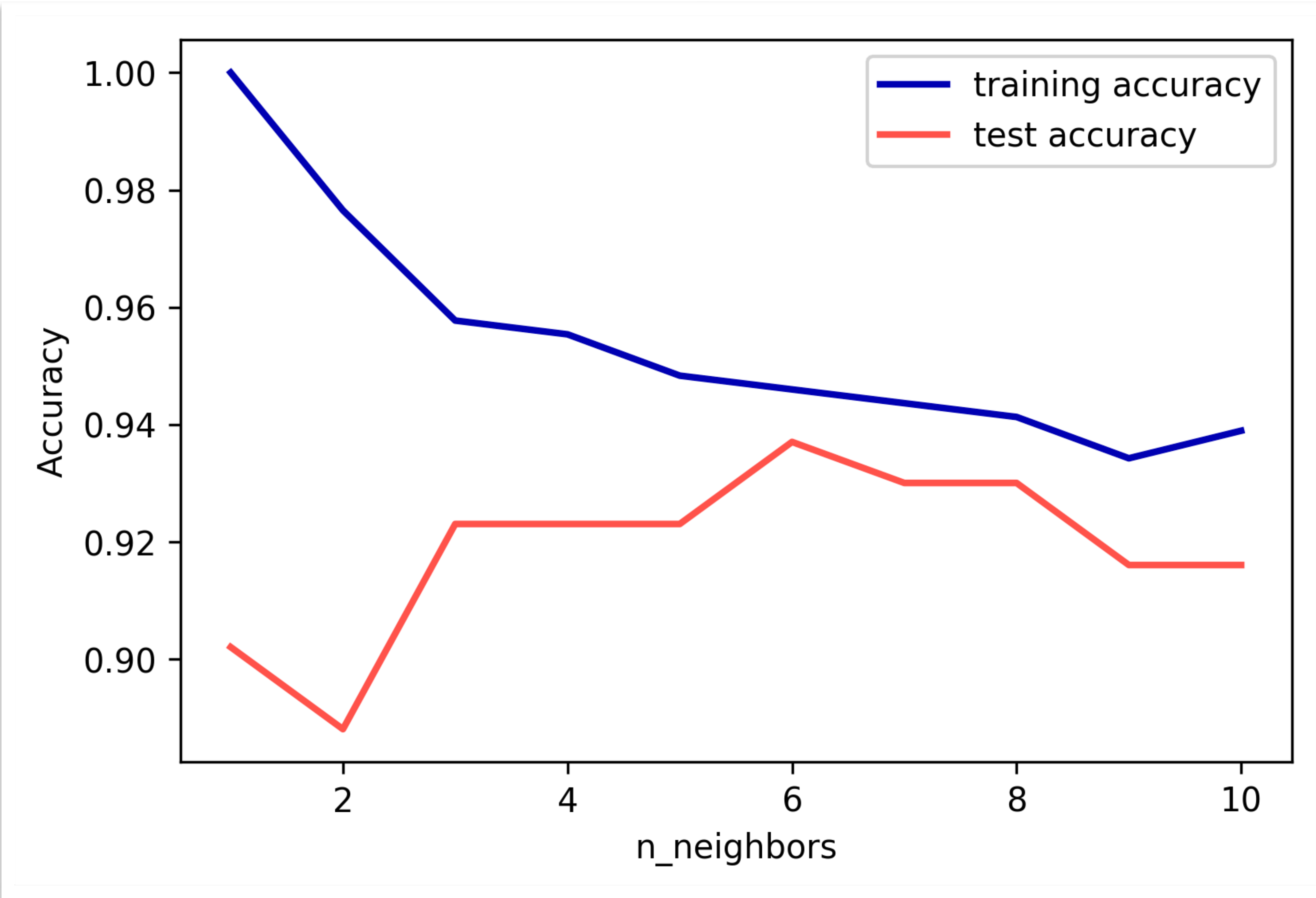
Puzzle of today!

$$\boxed{\text{Value of } K\text{?}}$$

the *K*

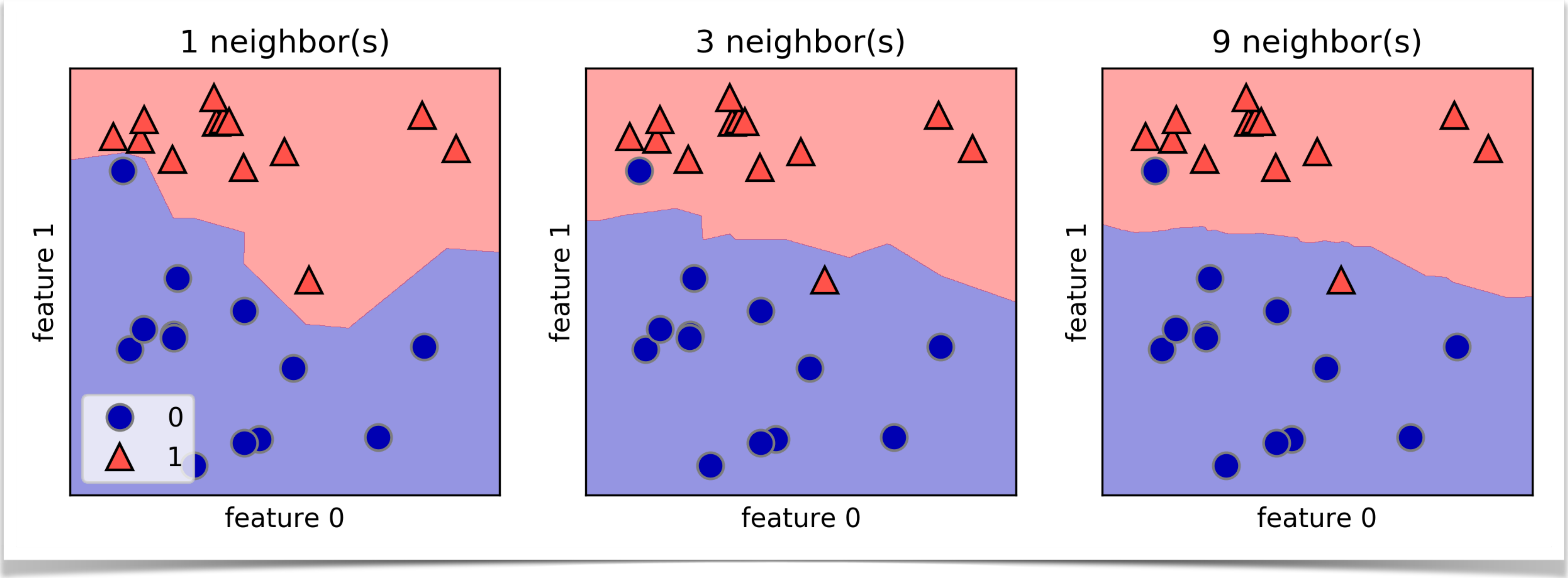
- Guessing
- Heuristics
  - use coprime class and K combination,
  - *K* is greater or equal to number of classes plus 1,
  - *K* that “avoids” noise (or outliers) —>> ***sqrt*(*n*)**
- Usually a tuned hyper-parameter.

Effect of *K*



training accuracy
test accuracy
Accuracy
1.00
0.98
0.96
0.94
0.92
0.90
2
4
6
8
10
n_neighbors

Effect of K



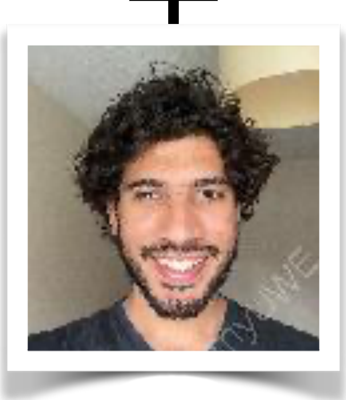
Example

observation	label
(3,3)	cool
(-1,-4)	not cool
(2,3)	cool
(0,-5)	not cool
…	…
…	…

Example

observation	label
(3,3)	cool
(-1,-4)	not cool
(2,3)	cool
(0,-5)	not cool

now given **Ali**, a new instance **(3,4)**, is he  ?



1st step: → choose ***K*** = **3**

2nd step: → compute ***distance***

data	***distance to (3,4)***
(3,3)	**1**
(-1,-4)	**8.9**
(2,3)	**1.4**
(0,-5)	**9.4**

Example

1st step: choose ***K*** = 3

2nd step: compute ***distance***

data	***distance to (3,4)***
(3,3)	**1**
(-1,-4)	**8.9**
(2,3)	**1.4**
(0,-5)	**9.4**

3d step: rank based on *d*

data	***rank*** *(closeness)*
(3,3)	**1**
(2,3)	**2**
(-1,-4)	**3**
(0,-5)	**4**

Example

1st step: → choose ***K*** = 3

2nd step: → compute ***distance***

data	***distance to (3,4)***
(3,3)	**1**
(-1,-4)	**8.9**
(2,3)	**1.4**
(0,-5)	**9.4**

3d step: → rank based on ***d***

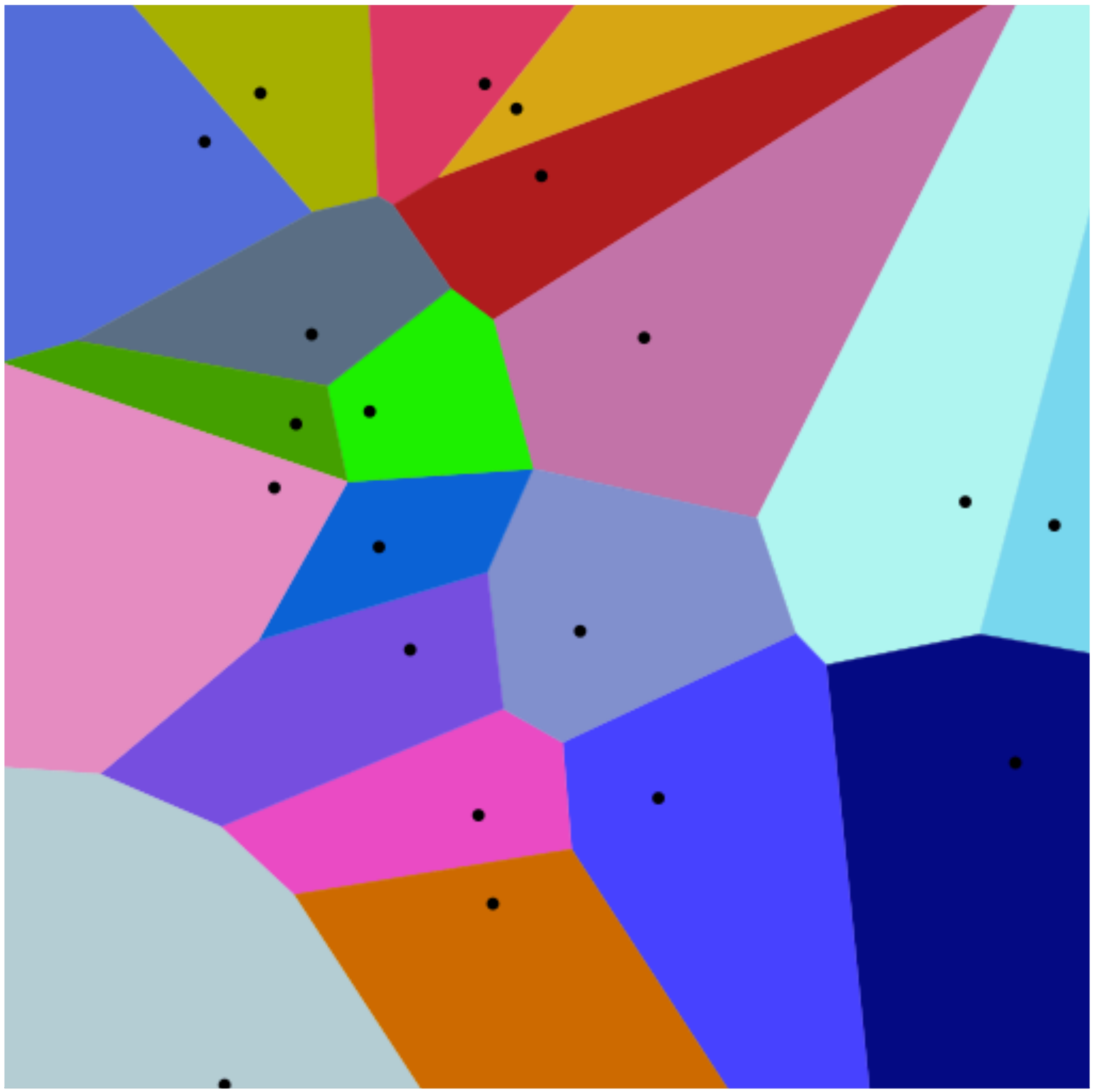
data	***rank*** *(closeness)*	
(3,3)	1	cool
(2,3)	2	cool
(-1,-4)	3	not cool
(0,-5)	~~4~~	

P(‘cool’ | **Ali**) = **2/3**

P(‘not cool’ | **Ali**) = **1/3**

***K*NN, with K=3, suggests that <u>Ali</u> is <u>most likely</u> ‘cool’**

Visualisation

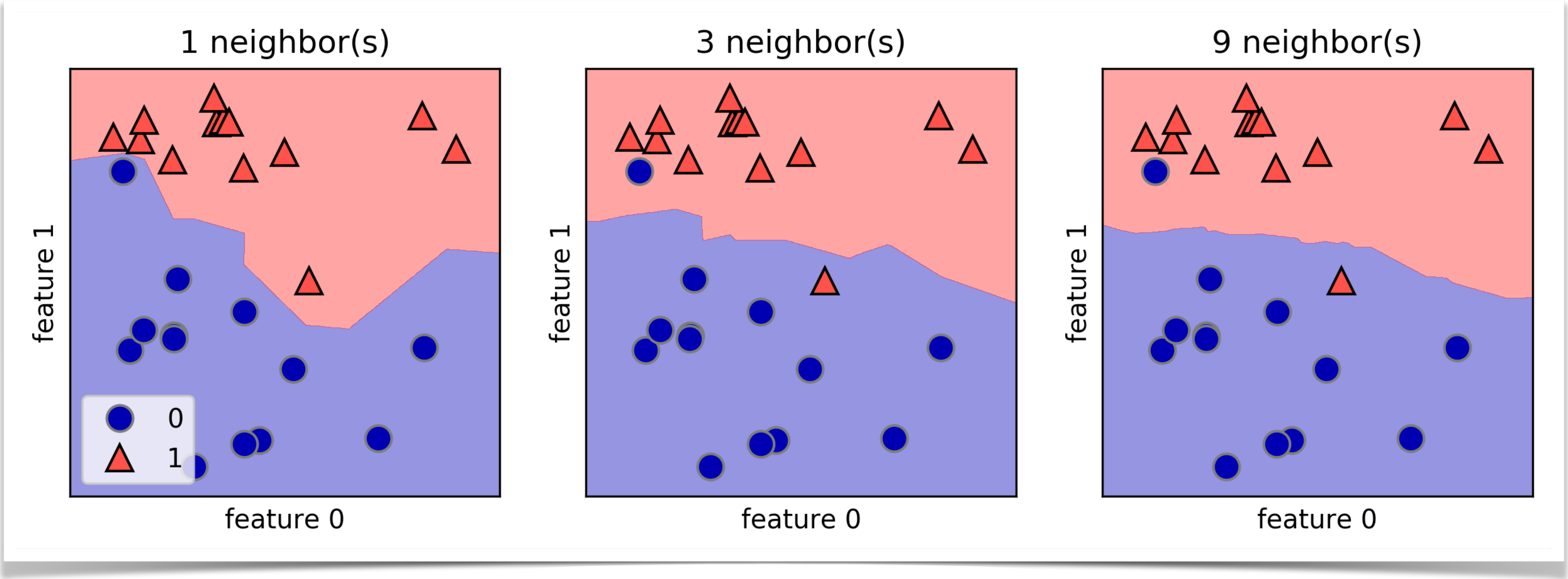


Georgy Voronoy

Born	28 April 1868 Zhuravki, Poltava Governorate, Russian Empire
Died	20 November 1908 (aged 40) Warsaw, Congress Poland
Citizenship	Russian Empire
Alma mater	Saint Petersburg University

Voronoi Diagram

Visualisation



Agenda

- ➡ K Nearest Neighbour/**Neighbor - kNN**
  - Properties
  - **Notes/thoughts**
  - Limitations
- ➡ Exercise - application of KNN for classification

Further notes/thoughts…

- Weighted voting: the inverse of ***d.***
- Distance metric choice; manhattan, euclidean, hamming, etc.
  - ➡ rescaling data: Scikit-learn's **normalize()**
- Dimensionality Reduction (PCA, Manifold, etc)
- Approximation of NN (k-d trees)

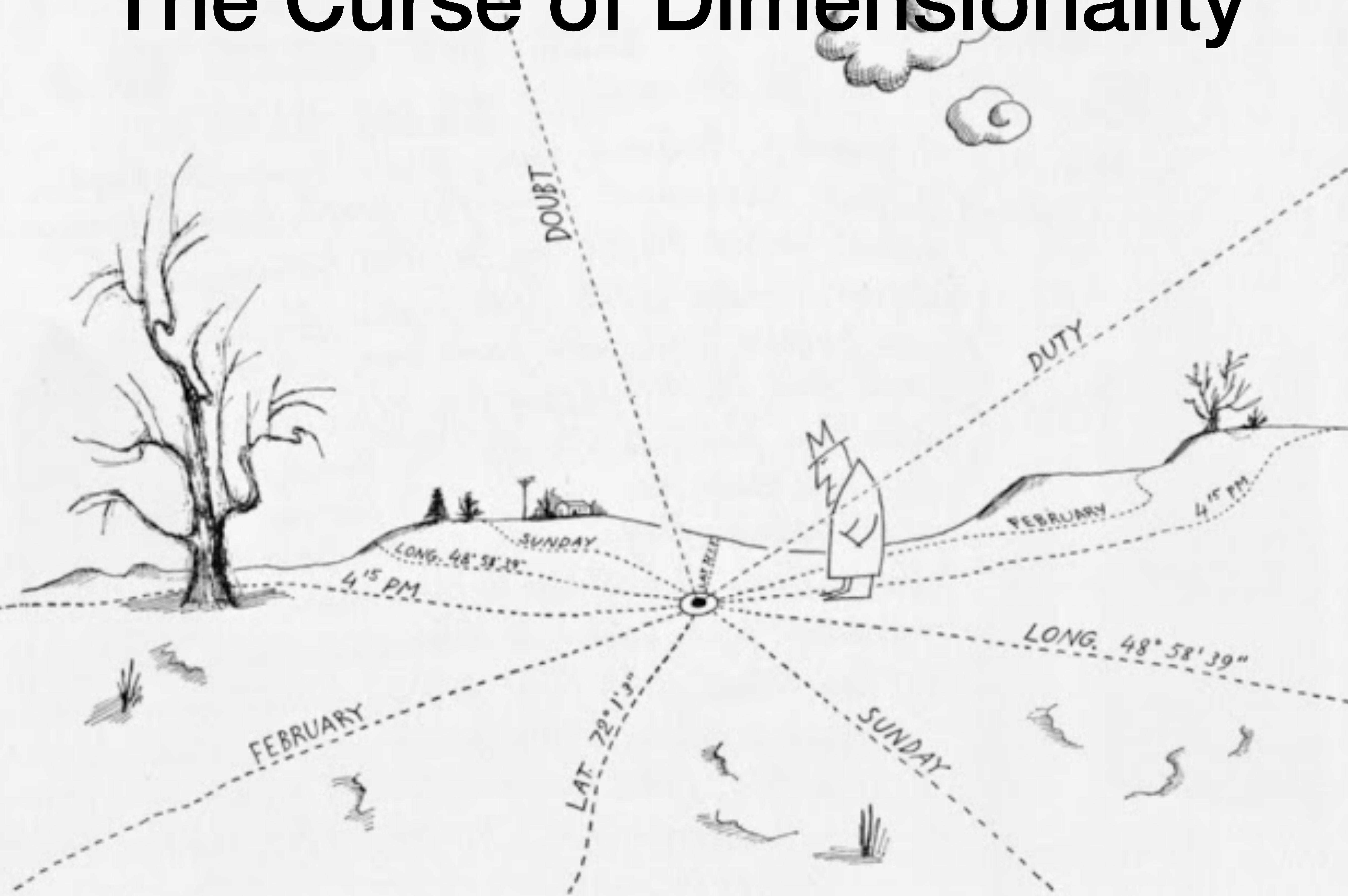
Agenda

➡ K Nearest Neighbour/**Neighbor - kNN**

- Properties
- Notes/thoughts
- **Limitations**

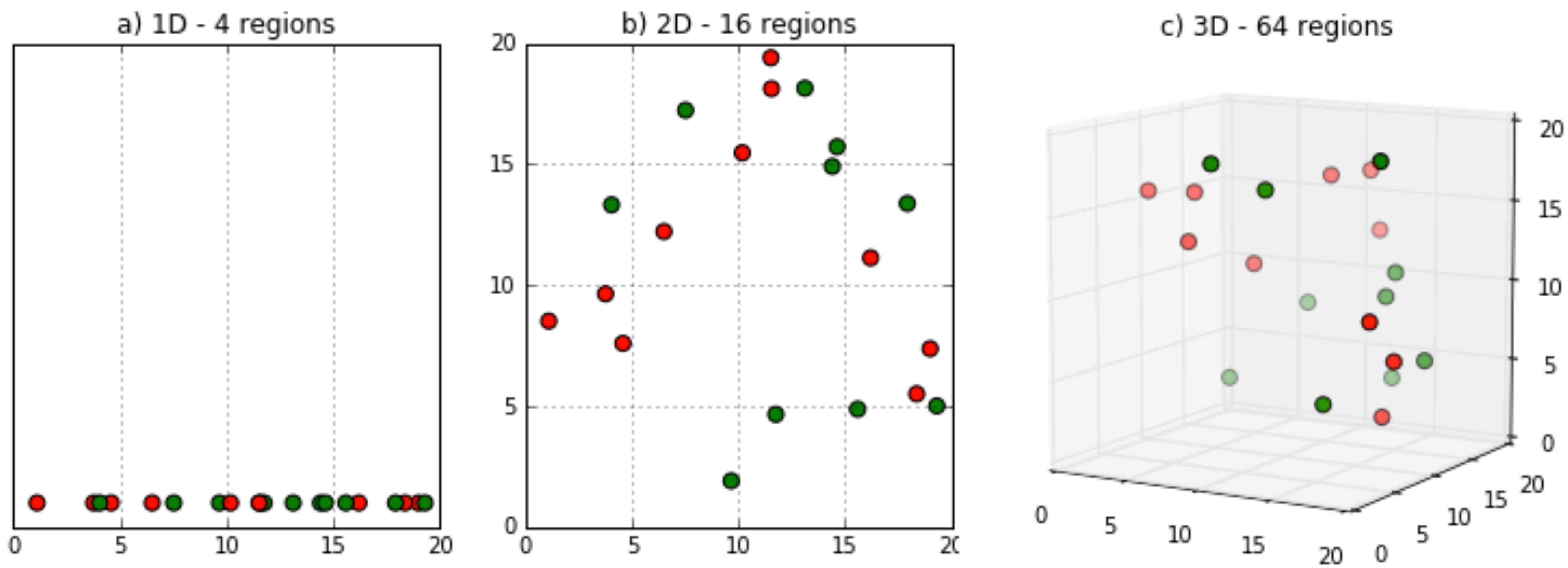
➡ Exercise - application of KNN for classification

The Curse of Dimensionality



Posted by Edward Tufte @EdwardTufte - The Curse of Dimensionality Saul Steinburg 1968

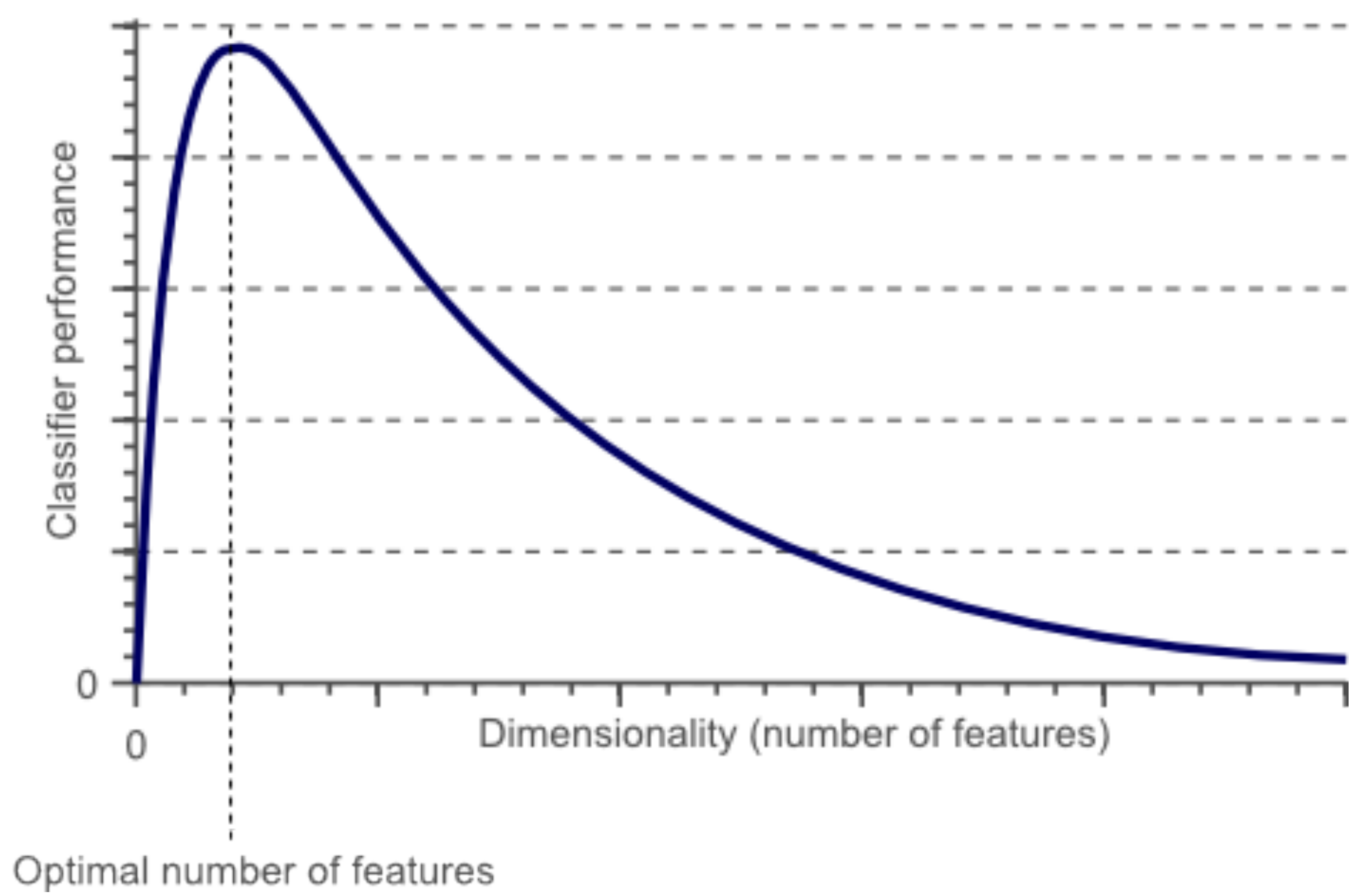
The Curse of Dimensionality



Recommended read: “The Curse of Dimensionality in classification” from Computer Vision for Dummies - available here.

Dimensions…

$$\lim_{d \to \infty} E\left(\frac{\text{dist}_{\max}(d) - \text{dist}_{\min}(d)}{\text{dist}_{\min}(d)}\right) \to 0$$



Applications

Proc AMIA Symp. 2000 : 759–763.

PMCID: PMC2243774

Application of K-nearest neighbors algorithm on breast cancer diagnosis problem.

M. Sarkar and T. Y. Leong

Author information ▶ Copyright and License information ▶

Abstract

This paper addresses the Breast Cancer diagnosis problem as a pattern classification problem. Specifically, this problem is studied using the Wisconsin-Madison Breast Cancer data set. The K-nearest neighbors algorithm is employed as the classifier. Conceptually and implementation-wise, the K-nearest neighbors algorithm is simpler than the other techniques that have been
Knearest neighbors algorithm produces the overall classifica
known for this problem.

Article link

Applied Mathematics in Electrical and Computer Engineering

Comparing Image Classification Methods: K-Nearest-Neighbor and Support-Vector-Machines

JINHO KIM[1]
Okemos High School
2800 Jolly Road
Okemos, MI 48864
[1]jcykim2008@gmail.com

BYUNG-SOO KIM[2], SILVIO SAVARESE[3]
Department of Electrical Engineering and Computer Science
University of Michigan
Ann Arbor, MI 48109-2122
[2]bsookim@umich.edu, [3]silvio@eecs.umich.edu

Abstract: - In order for a robot or a computer to perform tasks, it must recognize what it is looking at. Given an image a computer must be able to classify what the image represents. While this is a fairly simple task for

Article link

Applications

Introduction to

Information Retrieval

CS276: Information Retrieval and Web Search

Text Classification 1

Chris Manning and Pandu Nayak

Slides available here

k Nearest Neighbor Classification

- kNN = *k* Nearest Neighbor
- To classify a document *d*:
- Define *k*-neighborhood as the *k* nearest neighbors of *d*
- Pick the majority class label in the *k*-neighborhood
- For larger *k* can roughly estimate P(c|*d*) as #(c)/k

The bag of words representation

γ(I love this movie! It's sweet, but with satirical humor. The dialogue is great and the adventure scenes are fun… It manages to be whimsical and romantic while laughing at the conventions of the fairy tale genre. I would recommend it to just about anyone. I've seen it several times, and I'm always happy to see it again whenever I have a friend who hasn't seen it yet.)=c

Introduction to Information Retrieval

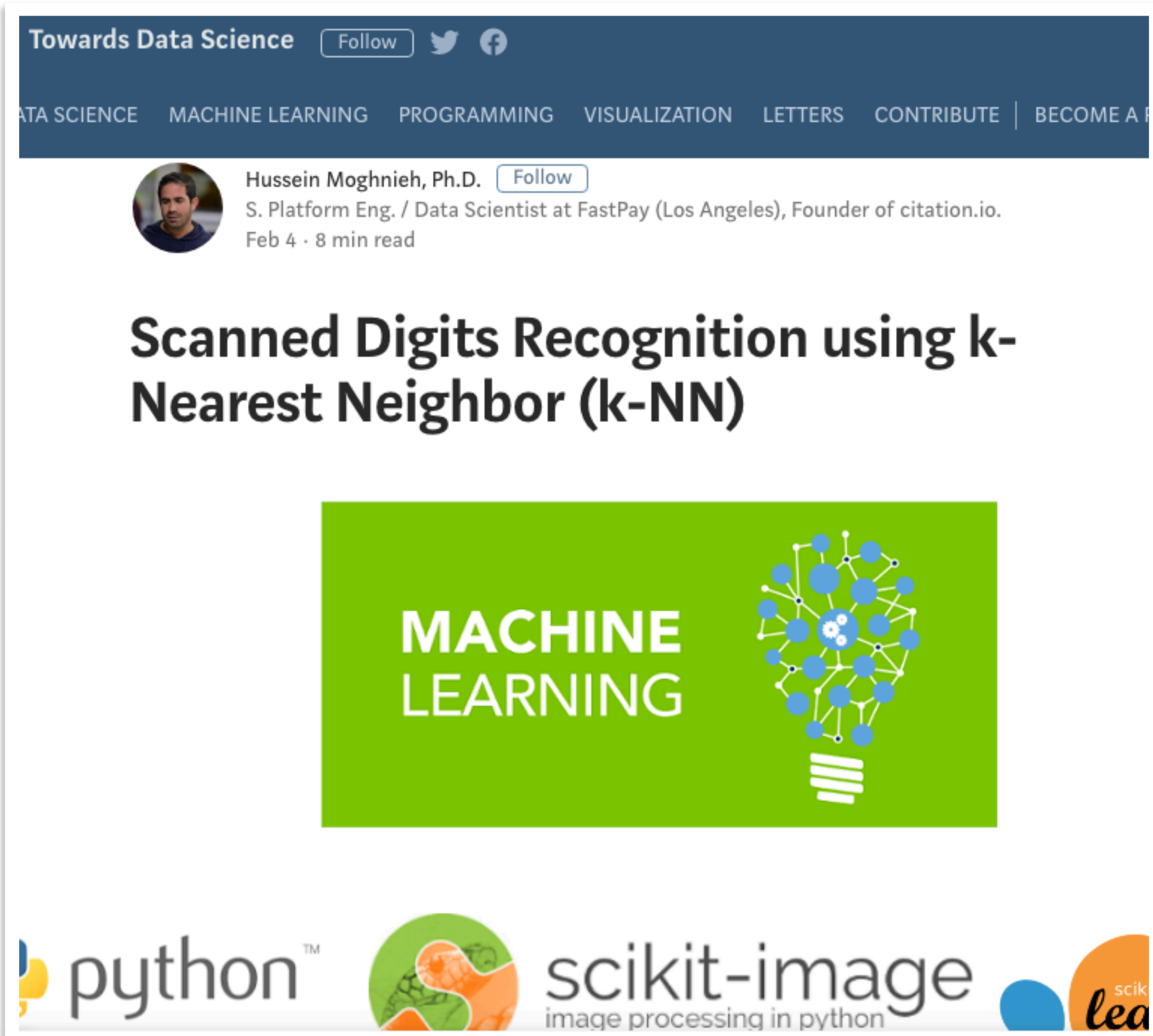
The bag of words representation

γ(

great	2
love	2
recommend	1
laugh	1
happy	1
...	...

)=c

Applications


Towards Data Science Follow

ATA SCIENCE MACHINE LEARNING PROGRAMMING VISUALIZATION LETTERS CONTRIBUTE | BECOME A

Hussein Moghnieh, Ph.D. Follow
S. Platform Eng. / Data Scientist at FastPay (Los Angeles), Founder of citation.io.
Feb 4 · 8 min read

Scanned Digits Recognition using k-Nearest Neighbor (k-NN)

[Page Link]()

Recap

➡ Simple, can be powerful, but lazy (and non-parametric).

➡ Commonly used in classification problems — it also is used for regression (which I didn't cover — reading provided on the latter).

➡ Works well on locally apparent patterns, on problems with low **intrinsic** feature space.

➡ Sensitive to feature scales.

➡ Though very commonly used, operationally isn't ideal.

Reading

- **J Grus (ML from Scratch):** Ch 12 (pp. 151–163)
- **E. Alpaydin (Intro ML 3d Ed):** Ch 8, Sec 8.2.3 (pp. 190–192)
- **T. Cover and P. Hart, "Nearest neighbor pattern classification," in *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21-27, January 1967.**
- **Read KNN Summary Sheet on BB/Week4 - do exercises if you have time.**