K-nearest neighbour

On Enhancing the performance of nearest neighbour classifiers using Hassanat distance metric,
Alkasassbeh, M., et al.

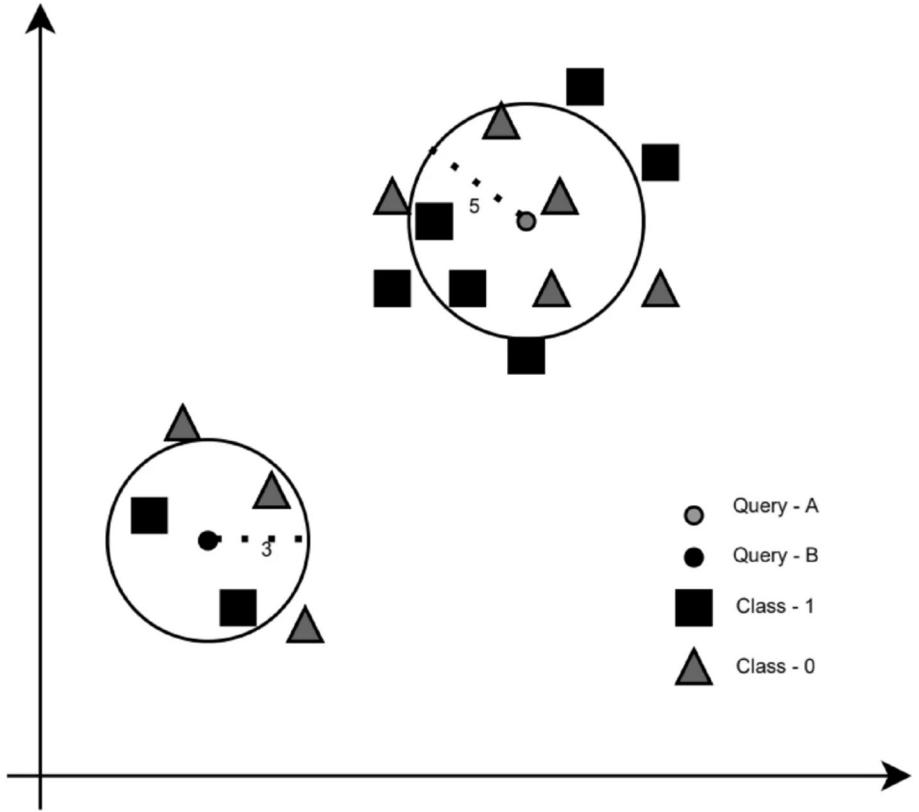
Presented by:
Noah Bloss
Frederik Hansen

Date: March 9, 2023

UNIVERSITY OF COPENHAGEN

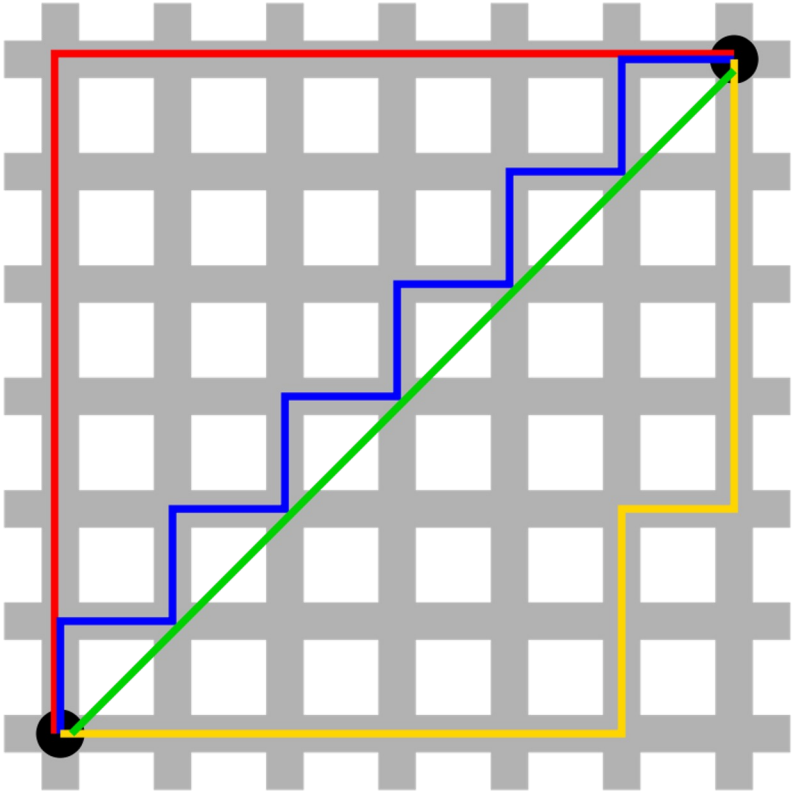
K-Nearest Neighbour

- Simple and effective classifier
- Choosing a class
  - Training the code
  - Classifying after training
- K as a variable
  - Most optimal K
  - Avoiding choosing a k



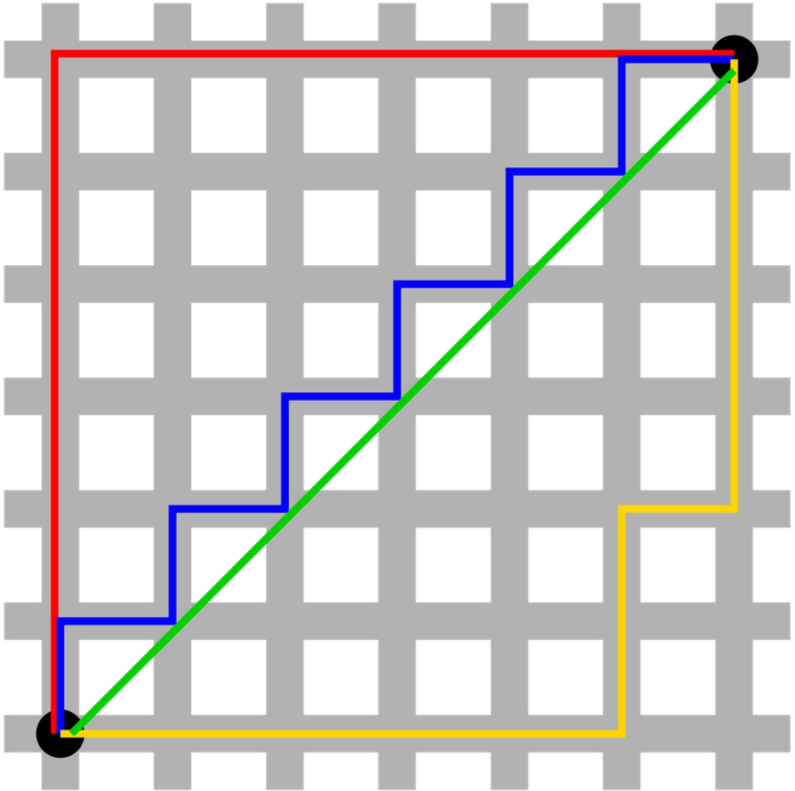
Distance Metrics

- Finding distance to nearest neighbour
- Common distance metrics
  - Euclidean
  - Manhattan



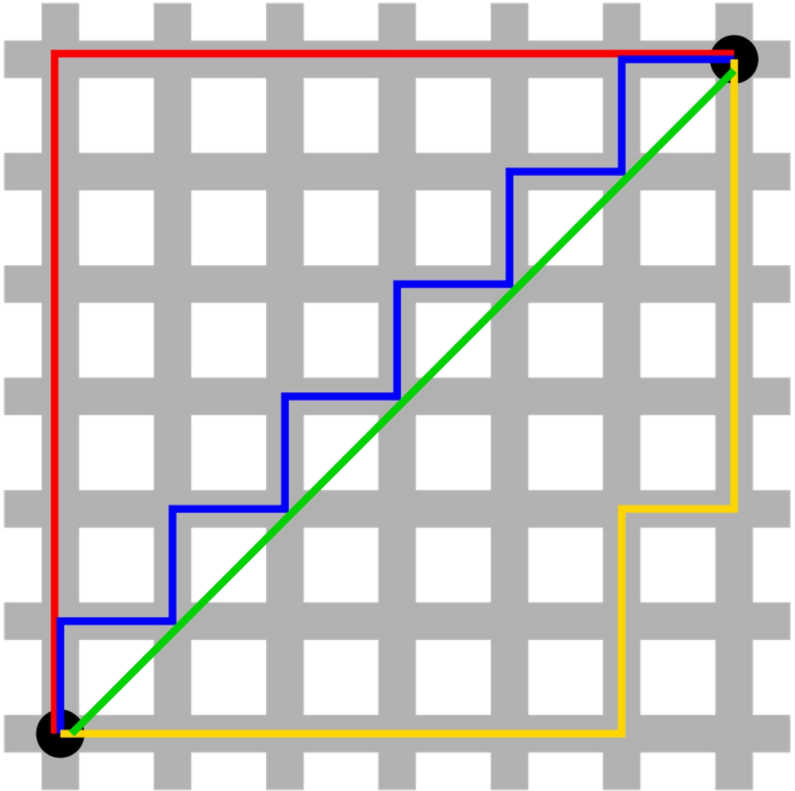
Distance Metrics

- Euclidean distance metric
  - $D_{Euclid}(x, x') = \sqrt{\sum_d |x_d - x'_d|^2}$
- Manhattan distance metric
  - $D_{Manhattan}(x, x') = \sum_d |x_d - x'_d|^2$



Distance Metrics

- Euclidean distance metric
  - $D_{Euclid}(x, x') = \sqrt{\sum_d |x_d - x'_d|^2}$
- Manhattan distance metric
  - $D_{Manhattan}(x, x') = \sum_d |x_d - x'_d|^2$
- Enhancing KNN with distance metrics
  - Hassanat distance metric



Distance Metrics - Hassanat

- Hassanat distance metric

$$D_{Hassanat}(A,B) = \sum_{i=1}^{m}(D(A_i,B_i))$$

$$D(A_i,B_i) = \begin{cases} 1 - \dfrac{1+\min(A_i,B_i)}{1+max(A_i,B_i)}, & min(A_i,B_i) \geq 0 \\ 1 - \dfrac{1+\min(A_i,B_i)+|\min(A_i,B_i)|}{1+max(A_i,B_i)+|\min(A_i,B_i)|}, & \min(A_i,B_i) < 0 \end{cases}$$

$$\lim_{\max(A_i,B_i)\to\infty}\big(D(A_i,B_i)\big) = \lim_{\min(A_i,B_i)\to-\infty}\big(D(A_i,B_i)\big) = 1$$

- Bounded in range [0,1[

Distance Metrics - Hassanat

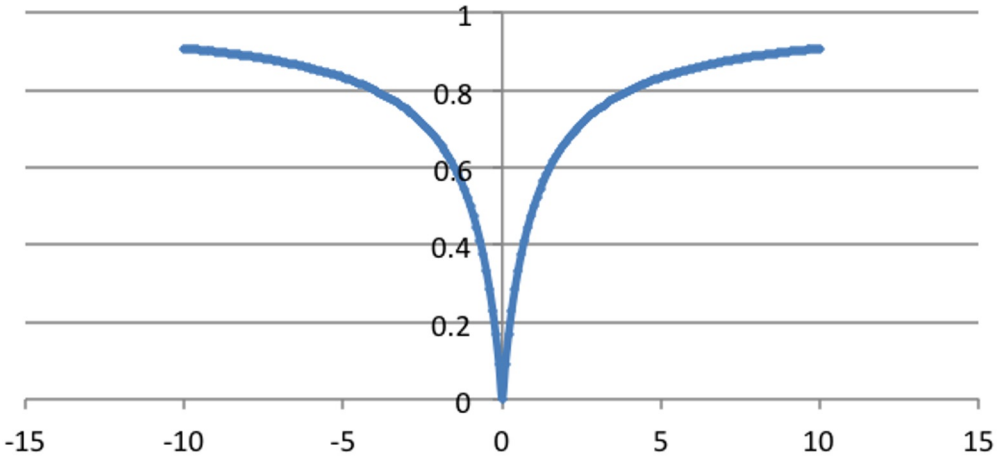
- Hassanat distance metric

$$D_{Hassanat}(A,B) = \sum_{i=1}^{m}(D(A_i,B_i))$$

$$\lim_{\max(A_i,B_i)\to\infty}\left(D(A_i,B_i)\right) = \lim_{min(A_i,B_i)\to-\infty}\left(D(A_i,B_i)\right) = 1$$

  - Bounded in range [0,1[

- Differences from other metrics
  - Not affected by noise
  - Outliers
  - Different data scale



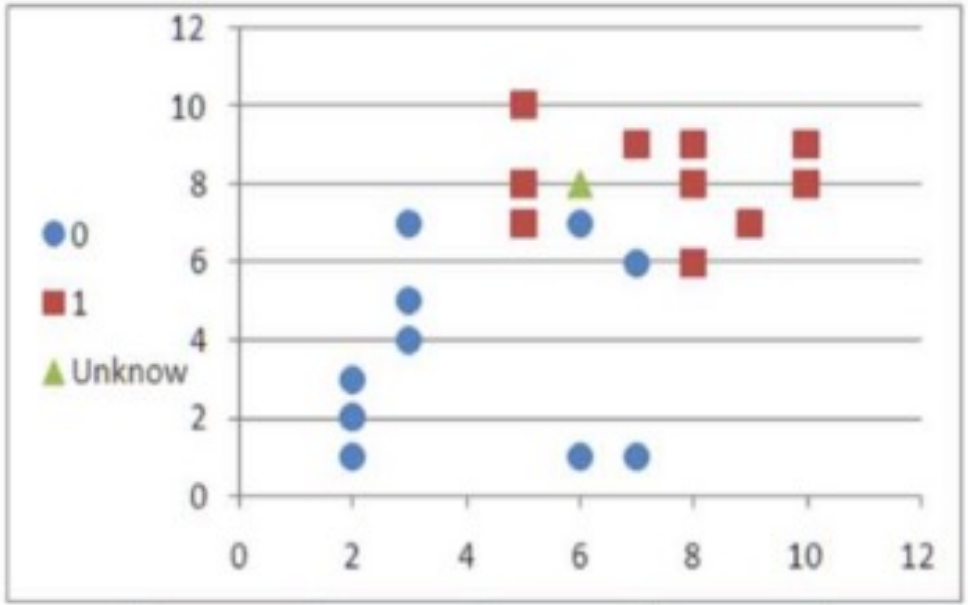
Inverted Indexes of Neighbours Classifier (IINC)

- K = number of training points
- Order distances – smallest first
- Summation of the inverted indexes:

$$S_c = \sum_{i=1(c)}^{L_c} \frac{1}{i}$$

- Influence is proportional to the distance
- Probability for each class:

$$P(x|c) = \frac{S_c}{S}, \quad S = \sum_{i=1}^{N} \frac{1}{i}$$



Ensemble Nearest Neighbour (ENN)

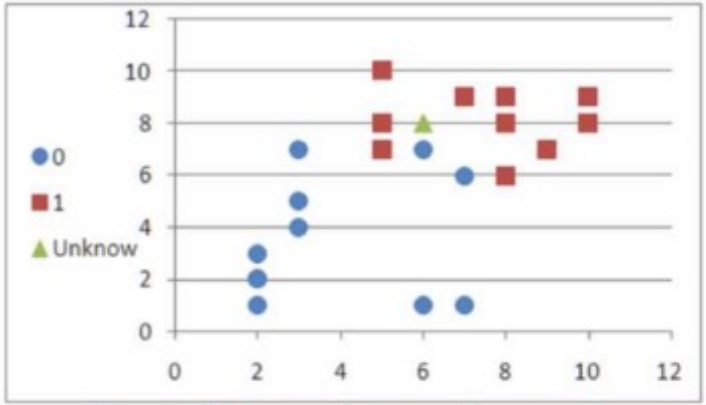
- K = all odd k up to $\sqrt{n}$
- Order distances – smallest first
- Weight of each point:

$$w(k) = \frac{1}{\log_2(1+k)}$$

- Shorter distance – more influence
- Weighted sum (WS)

$$WS_c = \sum_{k=1}^{\sqrt{n}} \sum_{i=1}^{k} \begin{cases} w(i), & A_i \\ 0, & Otherwise \end{cases}, \qquad k = k + 2$$

- $class = \underset{cWS_c}{argmax}$



class	0	1	1	1	1
order	1	2	3	4	5
weight	1	0.63	0.5	0.43	0.39
k	WS_0	WS_1			
1	1	0			
3	1	1.13			
5	1	1.95			
WS	3	3.08			

Result class=1

Comparison between enhanced and classic classifiers

- 28 data sets from the UCI Machine Learning Repository
  - 70% training data and 30% testing data
  - 10 different tests with training and test data random distributed
- Results
  - Comparing Manhattan and Hassanat distance metrics
  - Comparing classic KNN with enhanced (IINC and ENN)

Comparison between enhanced and classic classifiers

- Results using Manhattan distance metric

Algorithm	1NN	3NN	5NN	7NN	9NN	$\sqrt{n}$NN	IINC	ENN
Average	0.81	0.82	0.82	0.82	0.82	0.80	0.83	0.83

- Results using Hassanat distance metric

Algorithm	1NN	3NN	5NN	7NN	9NN	$\sqrt{n}$NN	IINC	ENN
Average	0.84	0.85	0.86	0.85	0.85	0.84	0.87	0.87

Comparison between enhanced and classic classifiers

- Increase in accuracy after applying Hassanat distance metric

Algorithm	1NN	3NN	5NN	7NN	9NN	$\sqrt{n}$NN	IINC	ENN
Average	0.03	0.03	0.04	0.04	0.04	0.04	0.03	0.03

Conclusion

- Improvement in performance
  - Hassanat distance metric improves performance
  - IINC and ENN improves performance
  - Data sets are not always similar → Classic KNN's are sometimes the best
- Missing comparison with Euclidean and other distance metrics
  - Better comparison → more trustworthy results

The end

Thank you for your attention!

Extra Slides

All tables

Manhattan	1NN	3NN	5NN	7NN	9NN	$\sqrt{n}$NN	IINC	ENN
Average	0.81	0.82	0.82	0.82	0.82	0.80	0.83	0.83

Hassanat	1NN	3NN	5NN	7NN	9NN	$\sqrt{n}$NN	IINC	ENN
Average	0.84	0.85	0.86	0.85	0.85	0.84	0.87	0.87

Difference	1NN	3NN	5NN	7NN	9NN	$\sqrt{n}$NN	IINC	ENN
Average	0.03	0.03	0.04	0.04	0.04	0.04	0.03	0.03