University of Nottingham
UK | CHINA | MALAYSIA

COMP3055 Machine Learning

Topic 6 – Instance Based Learning

Ying Weng

2024 Autumn

Instance Based Learning

- Directly compare new problem instances with instances seen in training
- No explicit modeling of the training data
- Complexity grows with the training data
- Classical instance based learning technique
  - **K Nearest Neighbour**

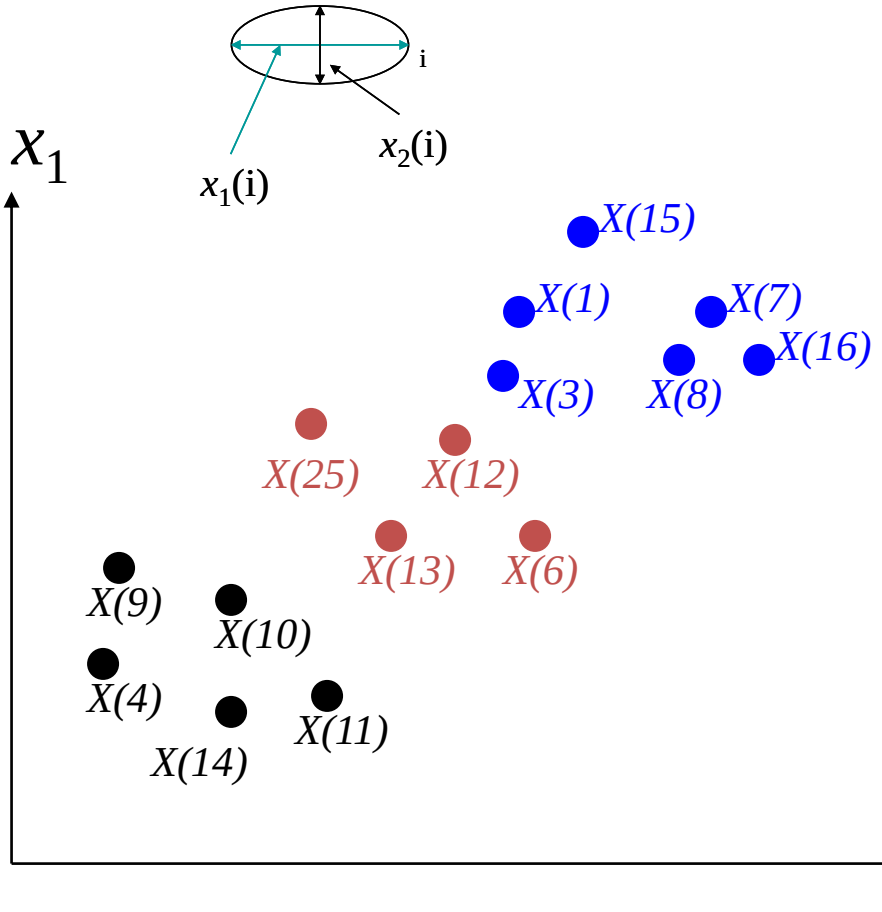
K-Nearest Neighbour Model

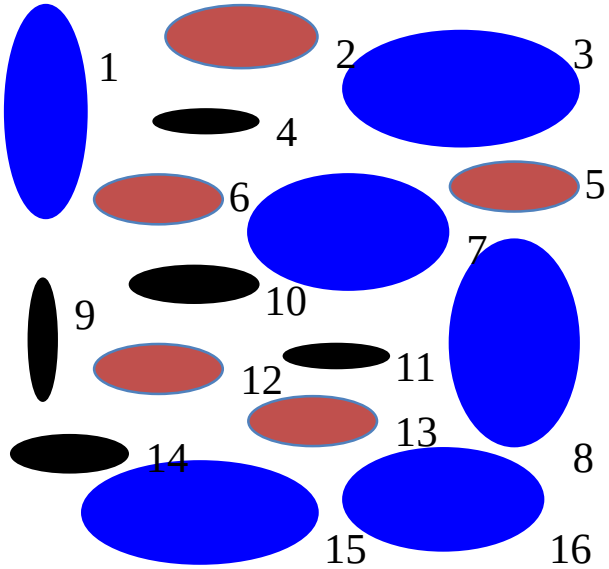
Example

Classify whether a customer will respond to a survey question using a 3-Nearest Neighbour classifier.

Customer	Age	Income	No. credit cards	Response
John	35	35K	3	No
Rachel	22	50K	2	Yes
Hannah	63	200K	1	No
Tom	59	170K	1	No
Nellie	25	40K	4	Yes
David	37	50K	2	?

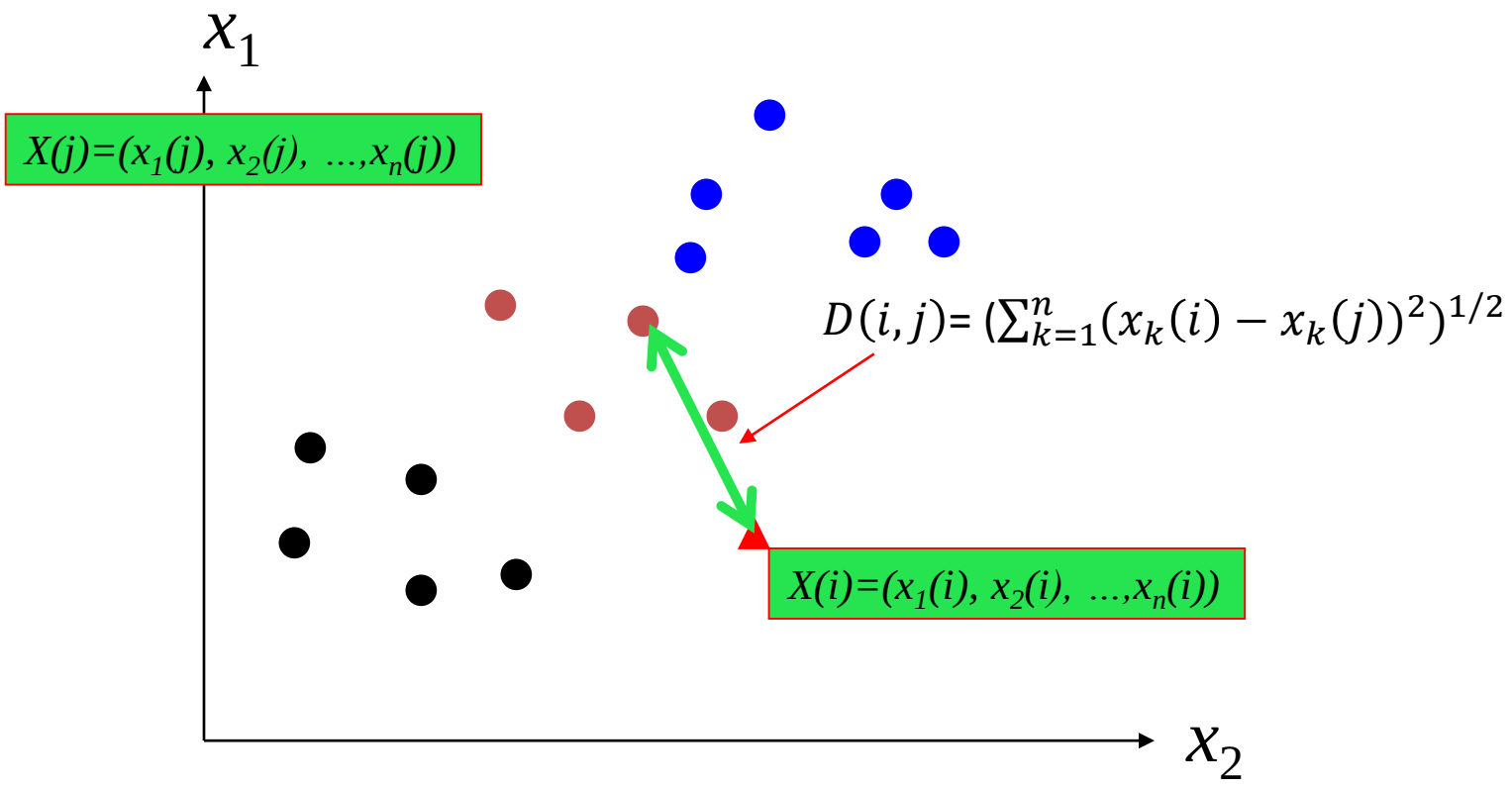
Objects, Feature Vectors, Points





Elliptical blobs (objects)

Nearest Neighbours



Nearest Neighbour Algorithm

Given training data *(X(1),D(1)), (X(2),D(2)), …, (X(N),D(N)),*

Define a distance metric between points in inputs space.
Common measures are:

Euclidean Distance

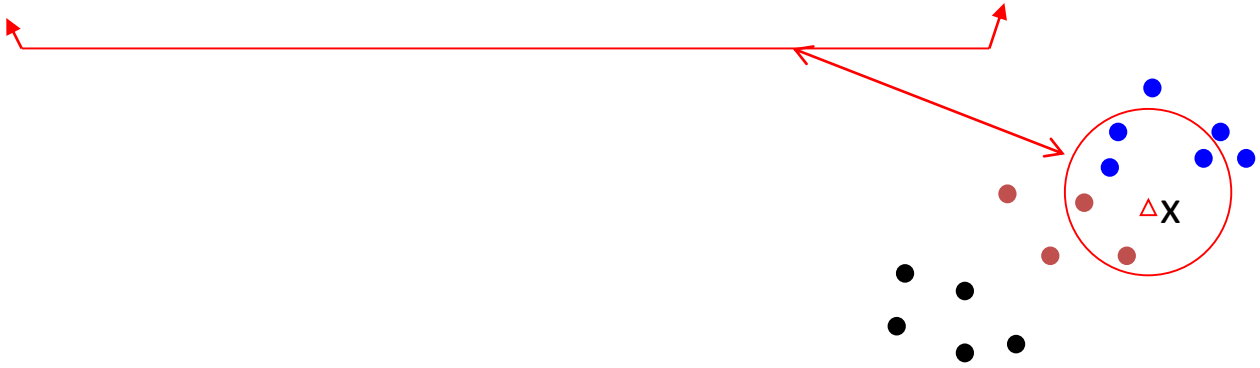
$D(i,j) = (\sum_{k=1}^{n}(x_k(i) - x_k(j))^2)^{1/2}$

K-Nearest Neighbour Model

Given test point X

- Find the K nearest training inputs to X
- Denote these points as

(X(1),D(1)), (X(2), D(2)), …, (X(k), D(k))



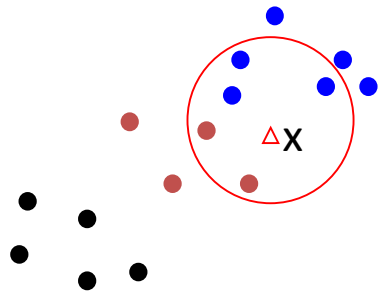
K-Nearest Neighbour Model

Instance based learning

The class identification of X

Y = most common class in set $\{D(1), D(2), \ldots, D(k)\}$

Majority rule



K-Nearest Neighbour Model

Example

Classify whether a customer will respond to a survey question using a 3-Nearest Neighbour classifier.

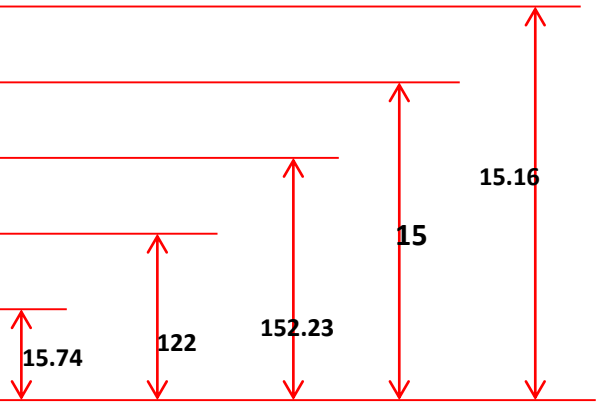
Customer	Age	Income	No. credit cards	Response
John	35	35K	3	No
Rachel	22	50K	2	Yes
Hannah	63	200K	1	No
Tom	59	170K	1	No
Nellie	25	40K	4	Yes
David	37	50K	2	?

K-Nearest Neighbour Model

Example

3-Nearest Neighbours

Customer	Age	Income	No. credit cards	Response
John	35	35K	3	No
Rachel	22	50K	2	Yes
Hannah	63	200K	1	No
Tom	59	170K	1	No
Nellie	25	40K	4	Yes
David	37	50K	2	?

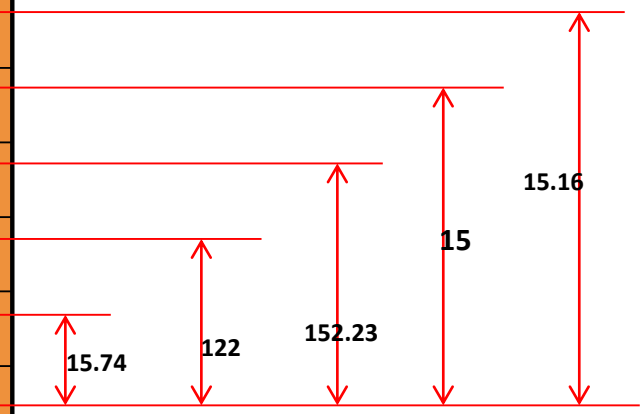


K-Nearest Neighbour Model

Example

3-Nearest Neighbours

Customer	Age	Income	No. credit cards	Response
John				No
Rachel				Yes
Hannah	63	200K	1	No
Tom	59	170K	1	No
Nellie				Yes
David	37	50K	2	?



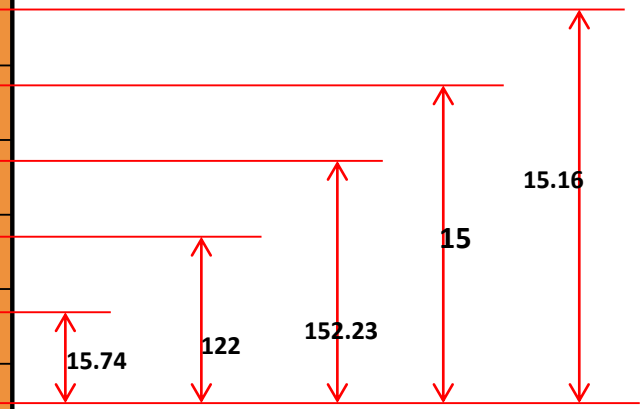
Three nearest ones to David are: No, Yes, Yes

K-Nearest Neighbour Model

Example

3-Nearest Neighbors

Customer	Age	Income	No. credit cards	Response
John				No
Rachel				Yes
Hannah	63	200K	1	No
Tom	59	170K	1	No
Nellie				Yes
David	37	50K	2	Yes



Three nearest ones to David are: No, Yes, Yes

K-Nearest Neighbour Model

Picking K

- Use *N fold cross validation* – Pick K to minimize the cross validation error

- For each of N training example

  - Find its K nearest neighbours
  - Make a classification based on these K neighbours
  - Calculate classification error
  - Output average error over all examples

- Use the K that gives lowest average error over the N training examples

K-Nearest Neighbour Model

Q&A

For the example we saw earlier, pick the best K from the set {1, 2, 3} to build a K-NN classifier.

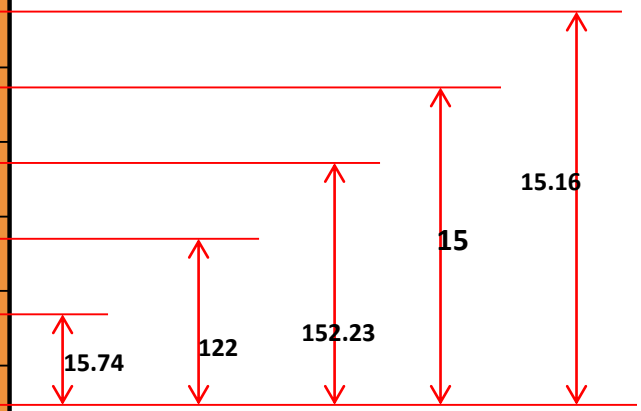
Customer	Age	Income	No. credit cards	Response
John	35	35K	3	No
Rachel	22	50K	2	Yes
Hannah	63	200K	1	No
Tom	59	170K	1	No
Nellie	25	40K	4	Yes
David	37	50K	2	?

K-Nearest Neighbour Model

Q&A

For the example we saw earlier, pick the best K from the set {1, 2, 3} to build a K-NN classifier.

Customer	Age	Income	No. credit cards	Response
John	35	35K	3	No
Rachel	22	50K	2	Yes
Hannah	63	200K	1	No
Tom	59	170K	1	No
Nellie	25	40K	4	Yes
David	37	50K	2	?



Further Readings

Chapter 8, T. M. Mitchell, Machine Learning, McGraw-Hill International Edition, 1997

Any Questions?