KTH ROYAL INSTITUTE OF TECHNOLOGY

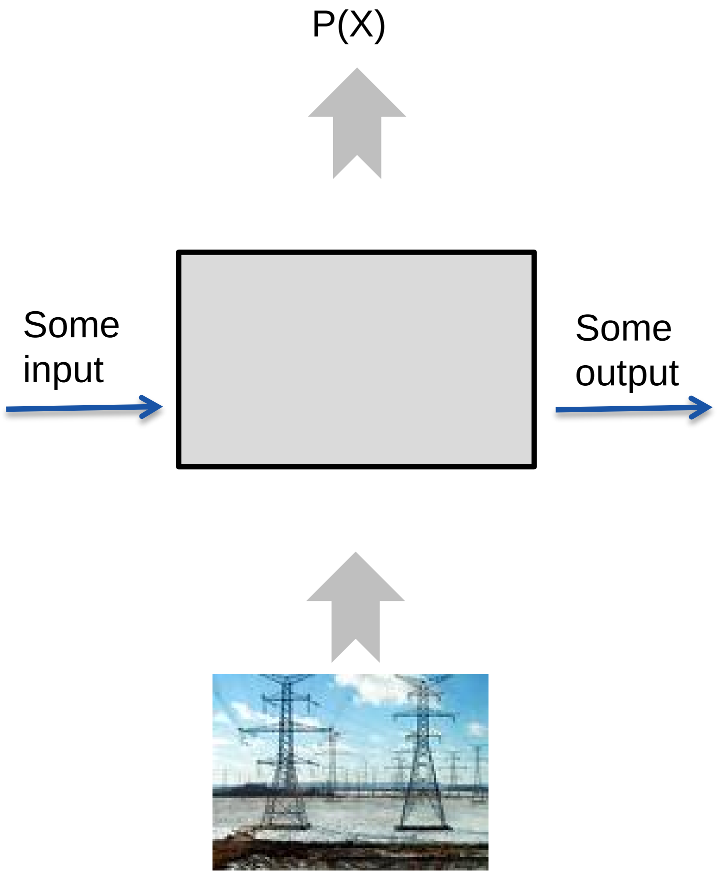
Lecture 14
Machine Learning. K-means, kNN

Contents

Power Systems Analysis –
An automated learning approach

Understanding states in the power system is established through observation of inputs and outputs without regard to the physical electrotechnical relations between the states.

Adding knowledge about the electrotechnical rules means adding heuristics to the learning.



Given a set of examples (the learning set (LS)) of associated input/output pairs, derive a general rule representing the underlying input/output relationship, which may be used to explain the observed pairs and/or predict output values for any new unseen input.

Supervised learning - a preview

In the scope of this course, we will be studying three forms of supervised learning.

- Decision Trees

  Overview and practical work on exercise session.

- Statistical methods – k-Nearest Neighbour

  Overview and practical work on exercise session. Also included in Project Assignment

  kNN algorithm can also be used for unsupervised clustering.

- Artificial Neural Networks

  Overview only, no practical work.

Lazy vs. Eager learning

In Eager learning, the Training set is pre-classified. All objects in the Learning set are clustered with regards to their neighbours. Ex.: Artificial Neural Networks.

In Lazy learning, only when a new object is input to the algorithm, the distance is calculated. Ex.: k Nearest Neighbours.

Contents

K-means clustering

K-Nearest Neighbour

K-means clustering

K-means clustering involves creating clusters of data

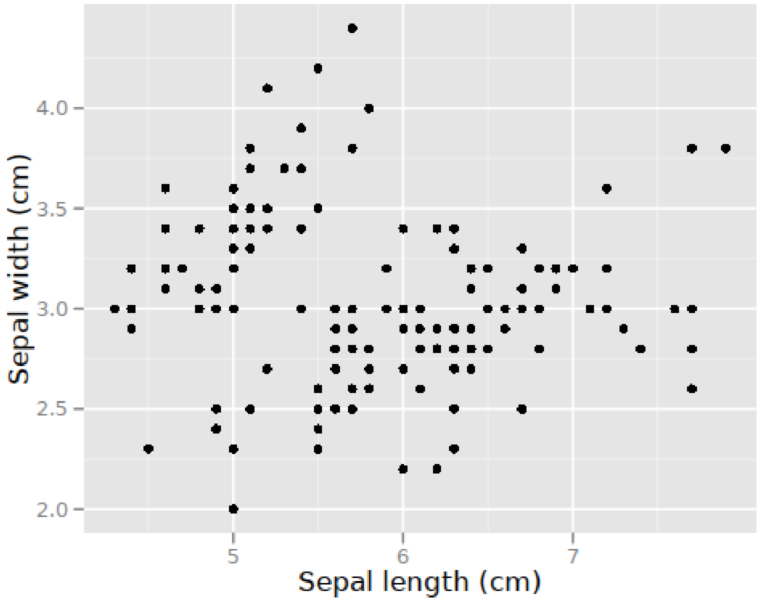
It is iterative and continues until no more clusters can be created

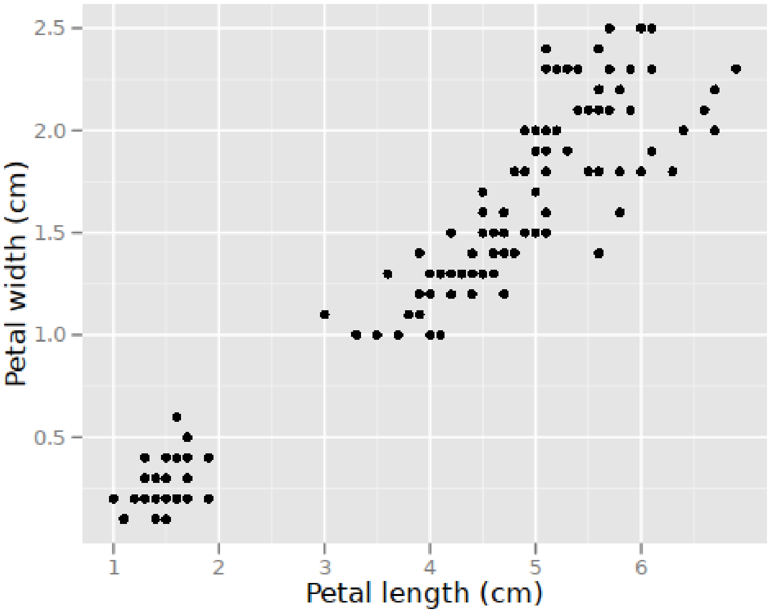
It requires the value of k to be defined at start.

Consider for instance a table like the following:

Sepal length	Sepal width	Petal length	Petal width
5.1	3.5	1.4	0.2
4.9	3.0	1.4	0.2
4.7	3.2	1.3	0.2

Plotted the data looks something like





K-means clustering (continued)

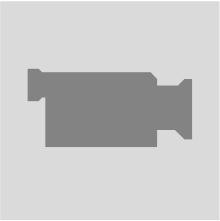
In k means clustering, first pick k mean points randomly in the space

Calculate the Euclidean distance from each data point in dataset to the mean points

Assign a data point to its closest mean point

Recalculate means

Once ended, we have k clusters



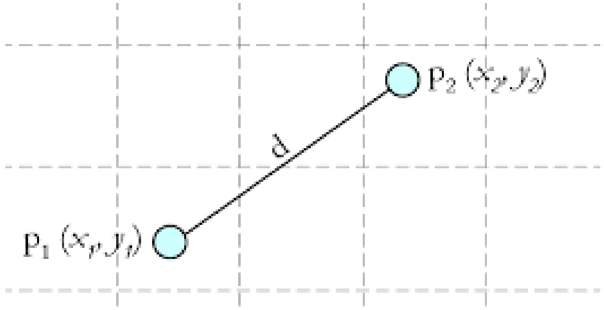
Contents

The k Nearest Neighbour algorithm

The k Nearest Neighbour algorithm is a way to classify objects with attributes to its nearest neighbour in the Learning set.

In kNN method, the **k** nearest neighbours are considered.

”Nearest” is measured as distance in Euclidean space.



Euclidean distance (d) $= \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$

k-Nearest Neighbour classification

Assuming instead a table like this where we have lables to ”clusters”

Sepal length	Sepal width	Petal length	Petal width	Species
5.1	3.5	1.4	0.2	iris setosa
4.9	3.0	1.4	0.2	iris setosa
4.7	3.2	1.3	0.2	iris setosa
…	…	…	…	...
7.0	3.2	4.7	1.4	iris versicolor
…	…	…	…	…
6.3	3.3	6.0	2.5	iris virginica
…	…	…	…	…

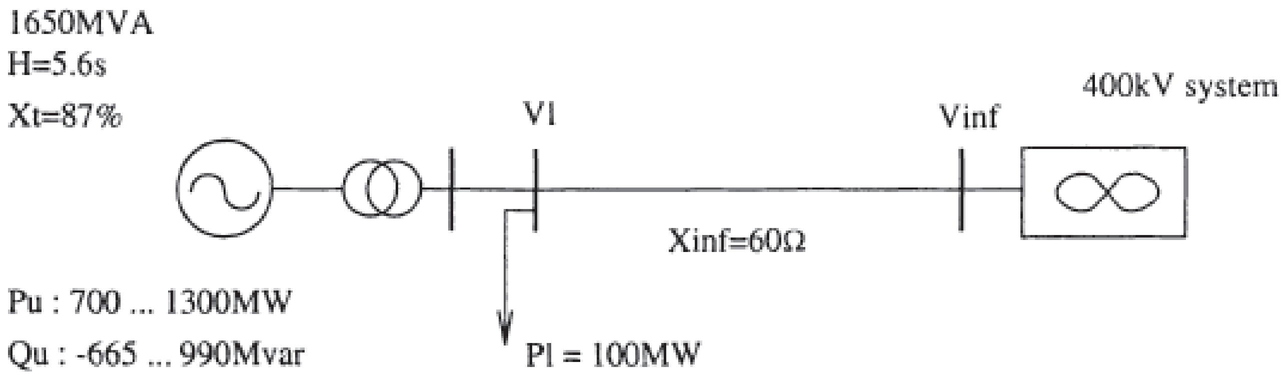
K-Nearest Neighbour algorithm

Given a new set of measurements, perform the following test:

Find (using Euclidean distance, for example), the k nearest entities from the training set. These entities have known labels. The choice of k is left to us.

Among these k entities, which label is most common? That is the label for the unknown entity.

Example from Automatic Learning techniques in Power Systems



One Machine Infinite Bus (OMIB) system

- Assuming a fault close to the Generator will be cleared within 155 ms by protection relays
- We need to identify situations in which this clearing time is sufficient and when it is not
- Under certain loading situations, 155 ms may be too slow.

Source: Automatic Learning techniques in Power Systems, L. Wehenkel

OMIB – further information

In the OMIB system the following parameters influence security

- Amount of active and reactive power of the generator (
- Amount of load nearby the generator (Pl)
- Voltage magnitudes at the load bus and at the infinite bus Short-circuit reactance Xinf, representing the effect of variable topology in the large system represented by the infinite bus.

In the example, Voltages at generator and Infinite bus are assumed similar and constant for simplicity

Source: Automatic Learning techniques in Power Systems, L. Wehenkel

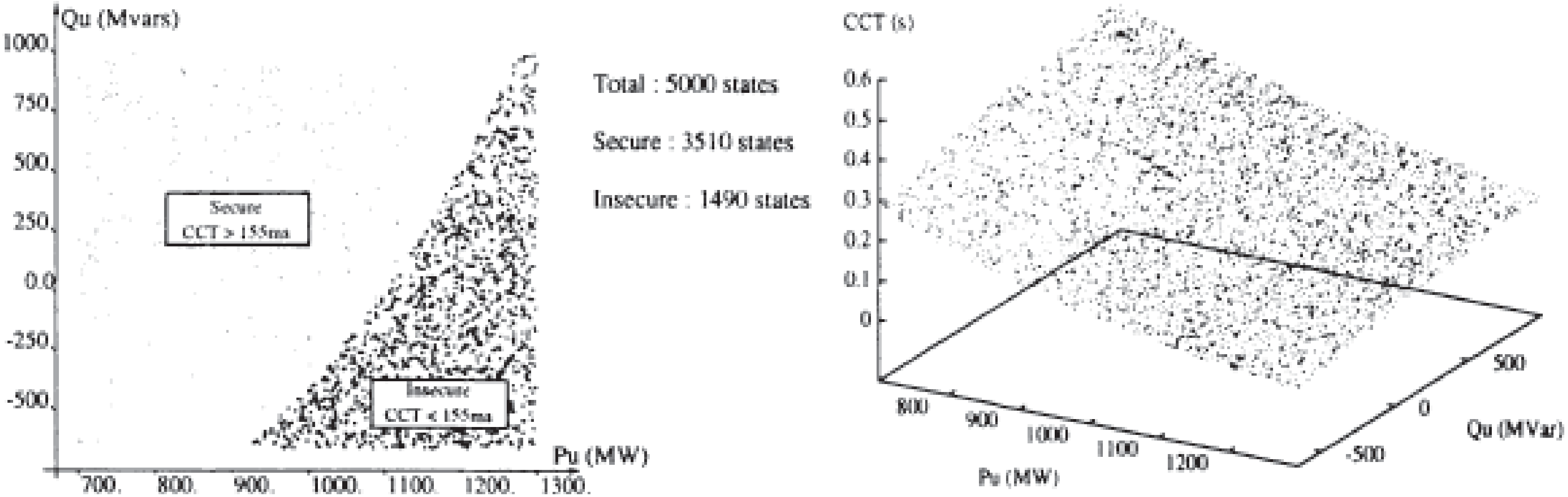
Our database of objects with attributes

In a simulator, we randomly sample values for P_u and Q_u creating a database with 5000 samples (objects) and for each object we have a set of attributes $(P_u, Q_u, V_1, P_1, V_{inf}, X_{inf}, CCT)$ as per below.

Table 1.1. Sample of OMIB operating states

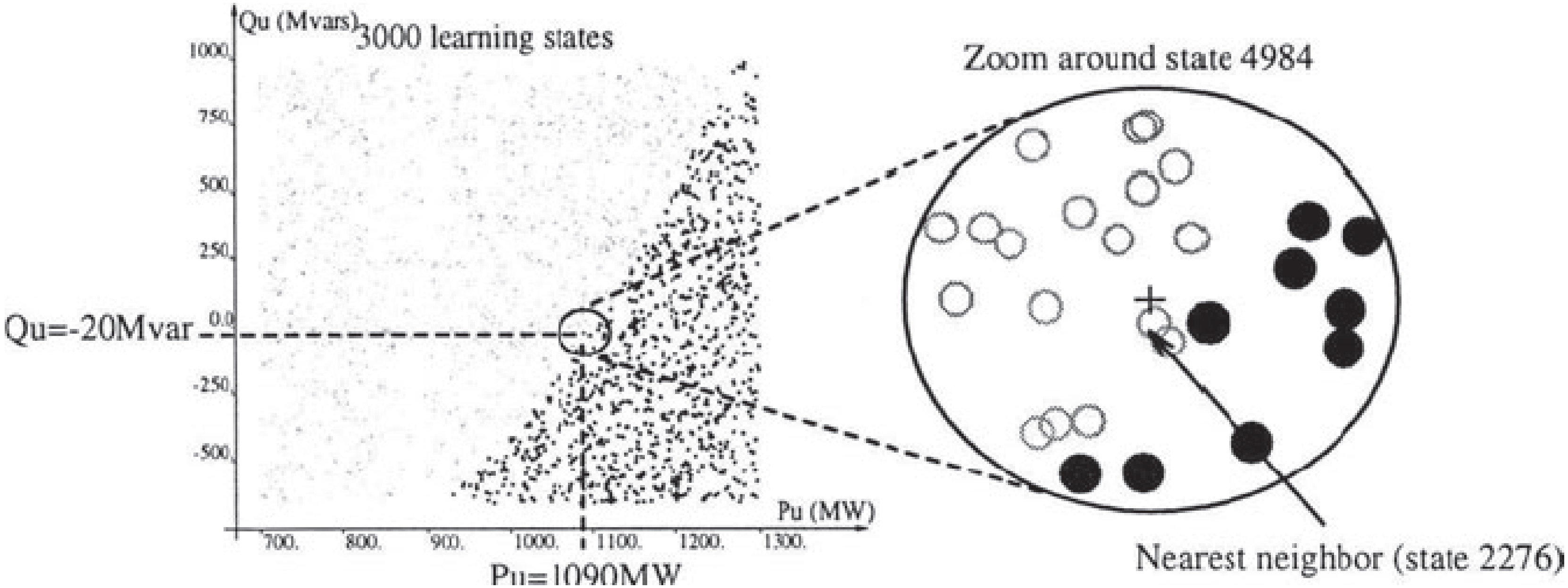
State Nb	Pu (MW)	Qu (MVAr)	V1 (p.u.)	P1 (MW)	Vinf (p.u.)	Xinf (Ω)	CCT (s)
1	876.0	-193.7	1.05	-100	1.05	60	0.236
2	1110.9	-423.2	1.05	-100	1.05	60	0.112
3	980.1	79.7	1.05	-100	1.05	60	0.210
4	974.1	217.1	1.05	-100	1.05	60	0.224
5	927.2	-618.5	1.05	-100	1.05	60	0.158
...	...	...	...	...	...	...	...
2276	1090.4	-31.3	1.05	-100	1.05	60	0.157
...	...	...	...	...	...	...	...
4984	1090.2	-20.0	1.05	-100	1.05	60	0.158
...	...	...	...	...	...	...	...

Plot of database content



In the OMIB example database

Sample 4984, and its neighbours



Error in the 1-NN classification