Data Mining and Synchrophasor Data

James Thorp

"1st International Symposium on Smart Grid Methods, Tools, and Technologies", Jinan, 18-20 May 2015.

VirginiaTech
Invent the Future

History of the use of Decision Trees in Power Systems

The combination of Decision Trees and PMUs was introduced in the ‘90s in connection with the installation of PMUs in the WECC in1994. [3]

At that point decision trees had been use in studying transient stability issues but had not been connected to the location of PMUs [5]. Decision trees generated by classification and regression trees (CART) have the useful ability to select the variables to be used in forming the tree from a longer list.

CART is a type of data mining software developed for broad application and is even used by cancer researchers in attempting to find cancer markers in DNA sequences.

In attempting to locate PMUs for a specific task in a large system CART proved to be invaluable.

The technique has been applied to data obtained from simulation for inputs to real-time, discrete-event control [4], predicting cascading events[15], voltage security[7], transient stability[9-10], detection of islanding[11], processing post disturbance records[12], security assessment[8], and adaptive security dependability of relays[14].

The availability of large amounts of real-time data has expanded the opportunities for these and other applications. OSIsoft is archiving WECC PMU data and has reported* a system for archiving 150,000 records a second.

*OSI T&D Users group meeting 2012

Outline

- History
- Example: Locating PMUs for adaptive security dependability [14]
  - Dealing with data as complex numbers
  - Higher dimensions: separating hyper planes
- Overview of Machine Learning in Power Systems
  - Classification and Regression
  - Clustering
  - Anomaly Detection
  - Bagging decision trees
  - Random Forest
  - Boosted Trees
  - Ensemble Trees BART

MapReduce
Open source
The Cloud

VirginiaTech
Invent the Future

Adaptive Security Dependability

- The protection system was designed to protect equipment – system was overbuilt. The system would work with a line out. If equipment was damaged the customers was out of service - high cost. Multiple primary (3 on transmission lines) protection and layers of backup protection. Backup of a backup
- A relay can do two things wrong – trip incorrectly or fail to trip. dependability is *"the degree of certainty that a relay or relay system will operate correctly"*,. Security *"relates to the degree of certainty that a relay or relay system will not operate incorrectly*
- The current system is dependable at the expense of security – trigger happy
- Adaptive Protection
- *Adaptive protection is a protection philosophy which permits and seeks to make adjustments automatically in in various protection functions in order to make them more attuned to prevailing system conditions*

Voting Scheme

- Adaptive Voting Scheme with three primary relays.

- State of the System
  - Stressed   Security = Vote
  - Safe   Dependability = Don't Vote

Adaptive Voting Scheme

- We are **NOT** changing relay settings neither during, before or after a fault.
- The choice of location for the measurements and the voting logic is obtained using data mining software.
- We create the data base by running many simulations (15,000) CART Classification and Regression Trees

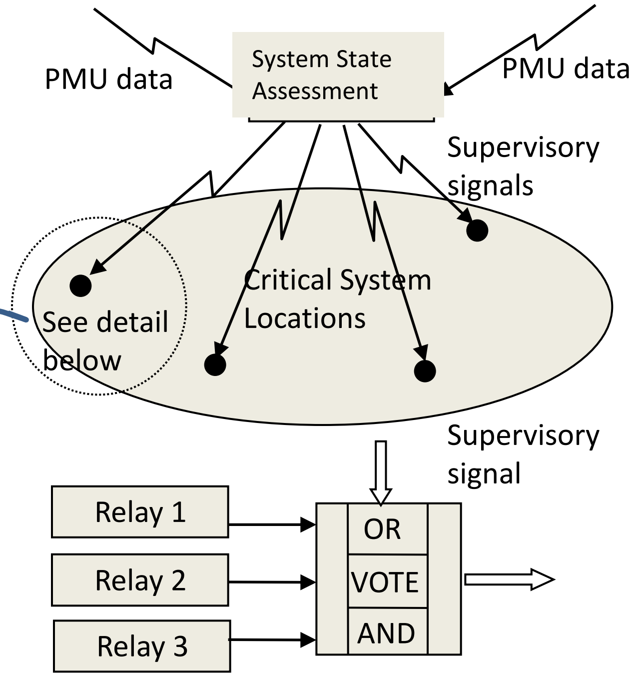
Adjusting balance of security-dependability

What terminal?

What measurements?

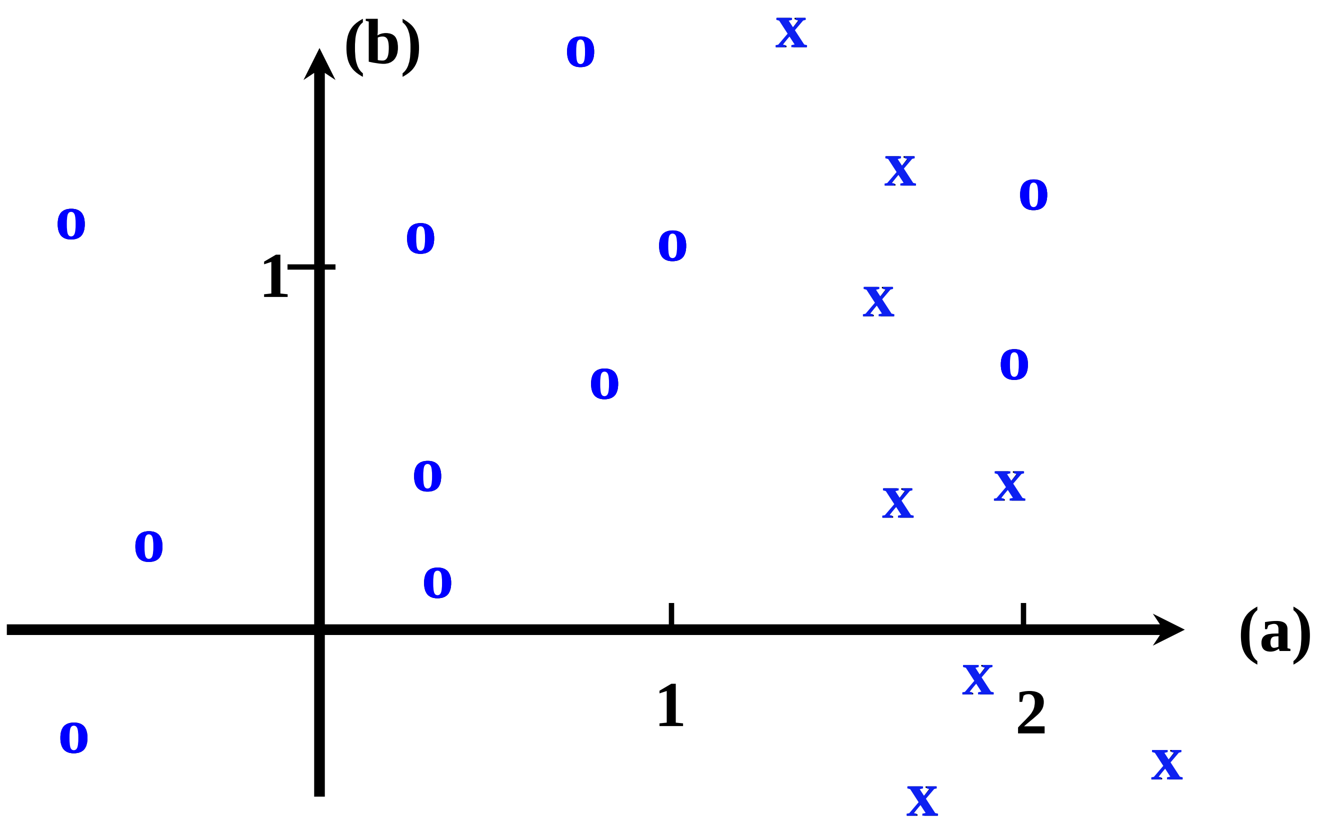
Determination of triggering logic

Performance evaluation



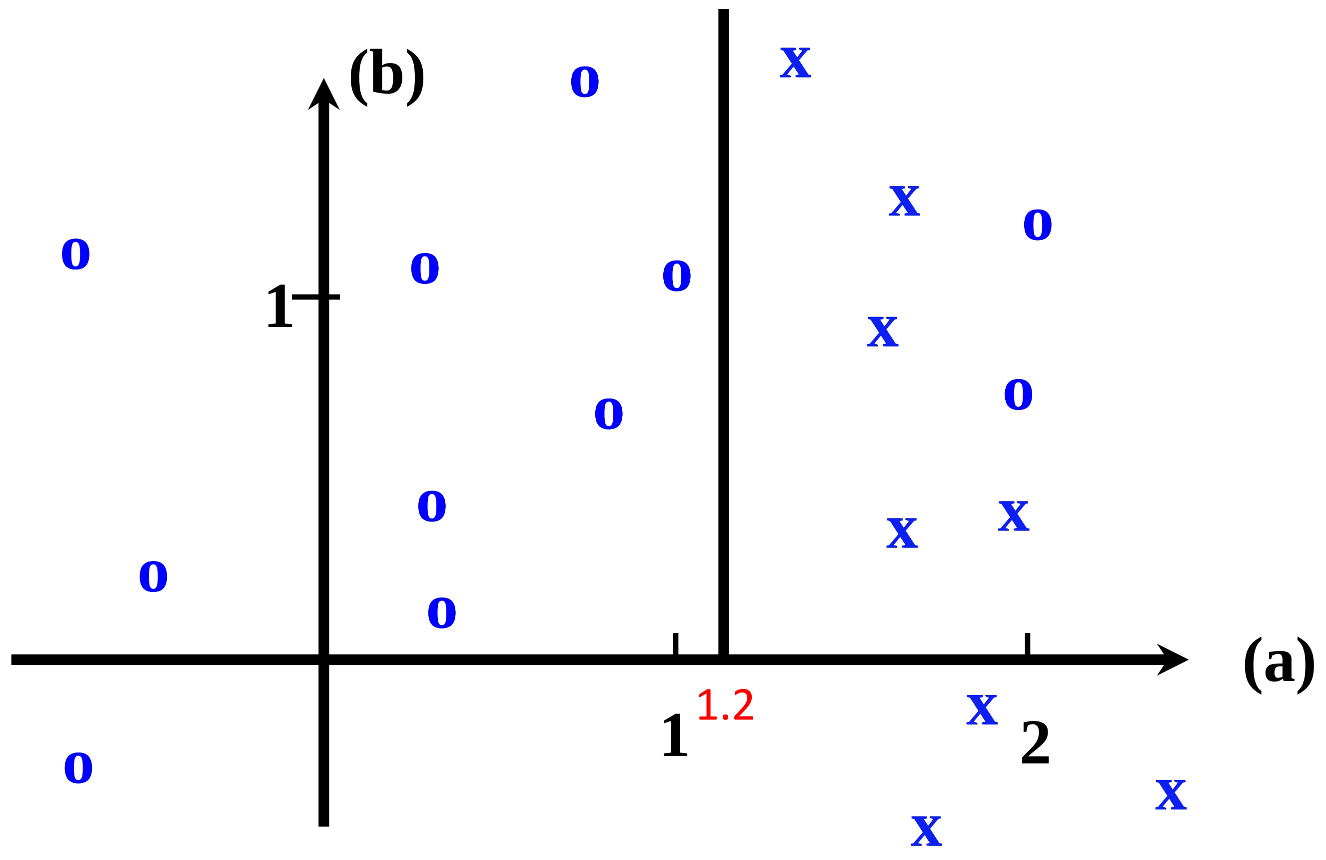
Adjustment of Dependability-Security balance under stressed system conditions.

Recursive Partitioning Algorithm
goal separate xs and os



CART selects splitting variables and the logic of the tree. That is, CART selects PMU locations and the logic.

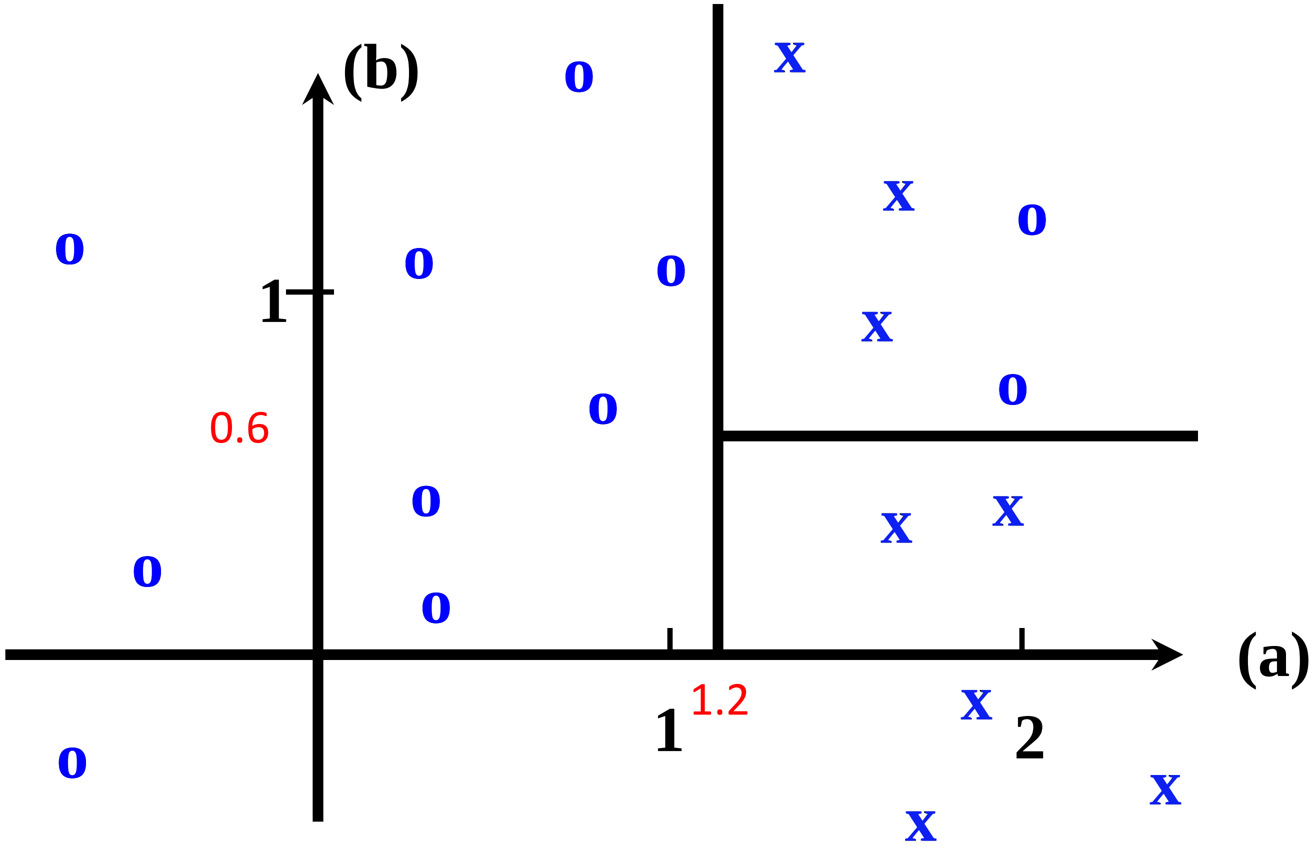
Recursive Partitioning Algorithm first



a < 1.2 done

CART selects splitting variables and the logic of the tree. That is, CART selects PMU locations and the logic.

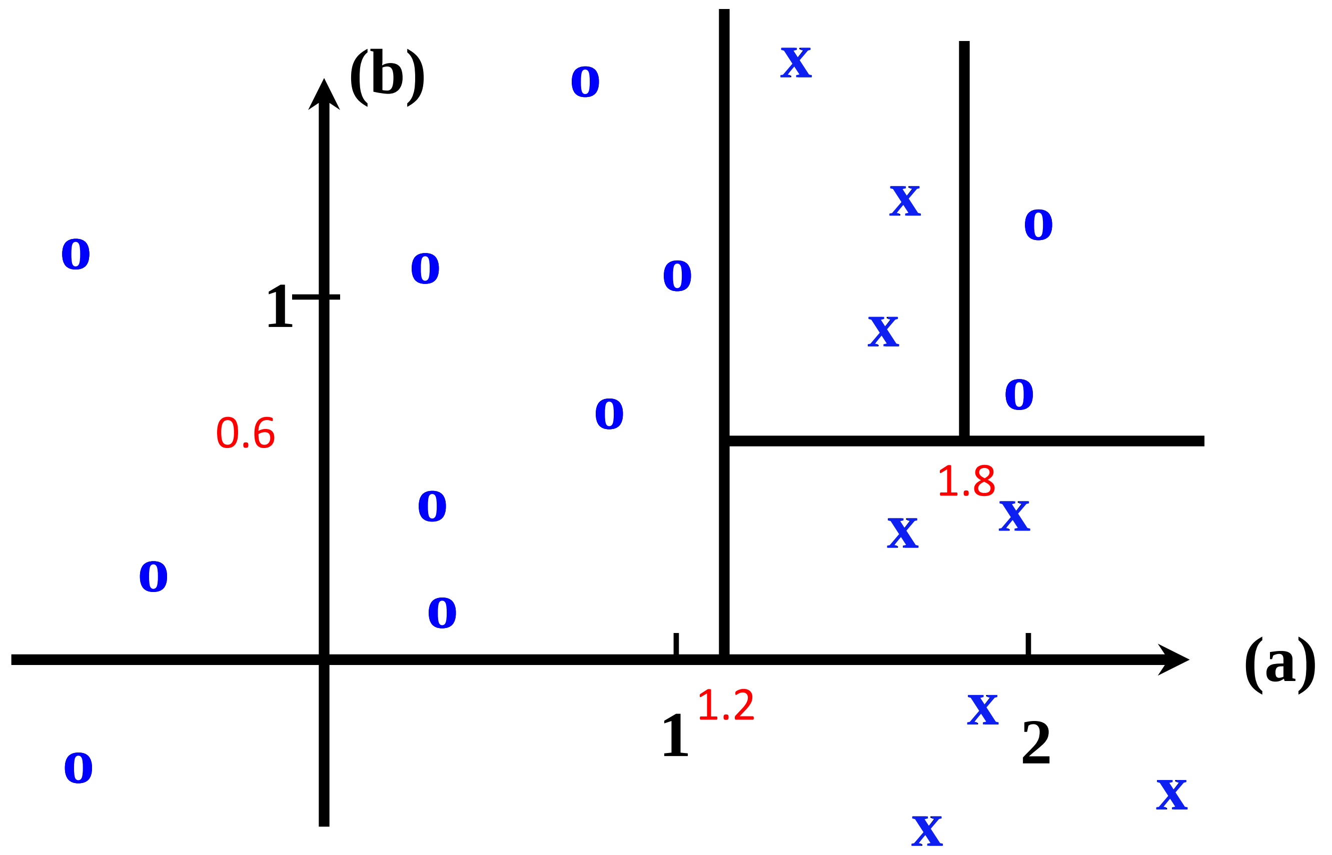
Recursive Partitioning Algorithm second



a< 1.2
b<.6 done

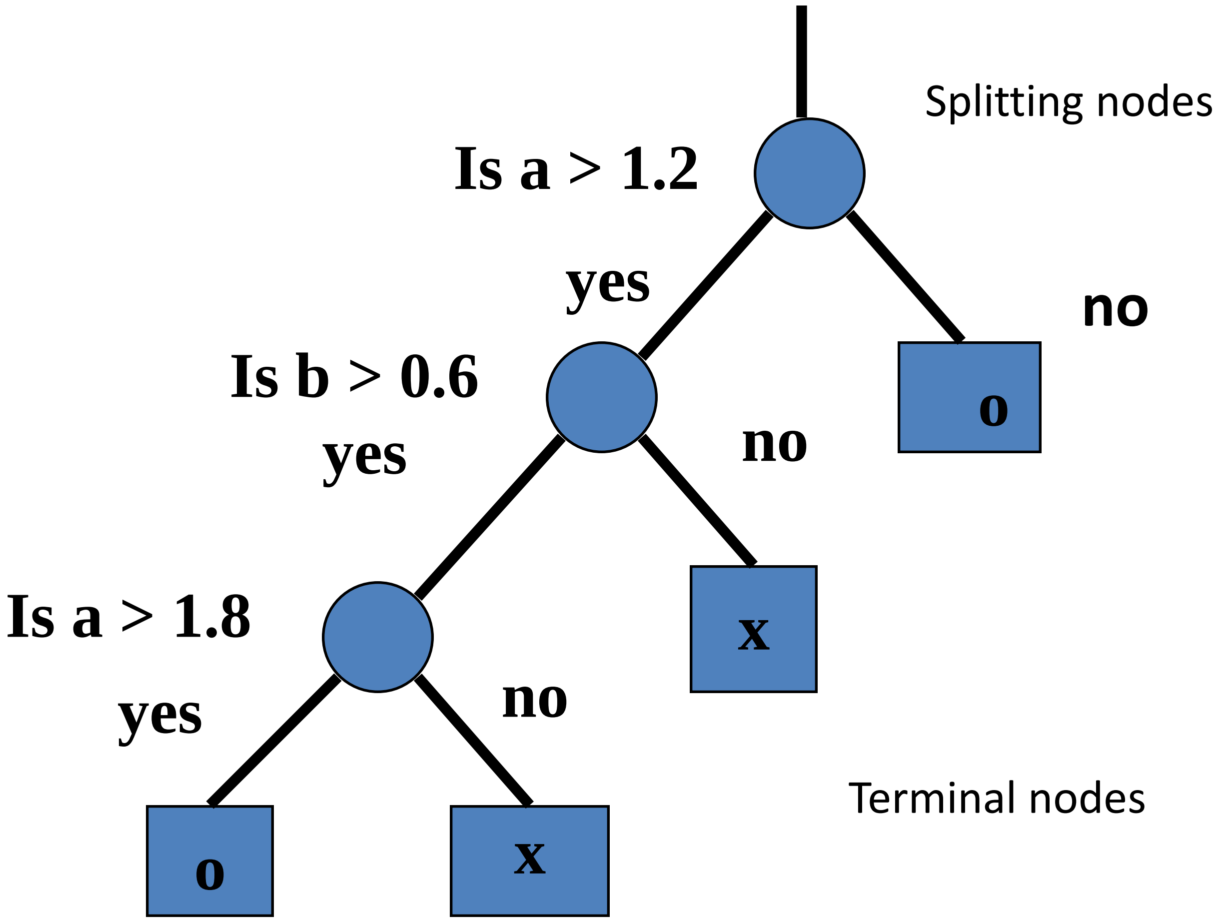
CART selects splitting variables and the logic of the tree. That is, CART selects PMU locations and the logic.

Recursive Partitioning Algorithm third



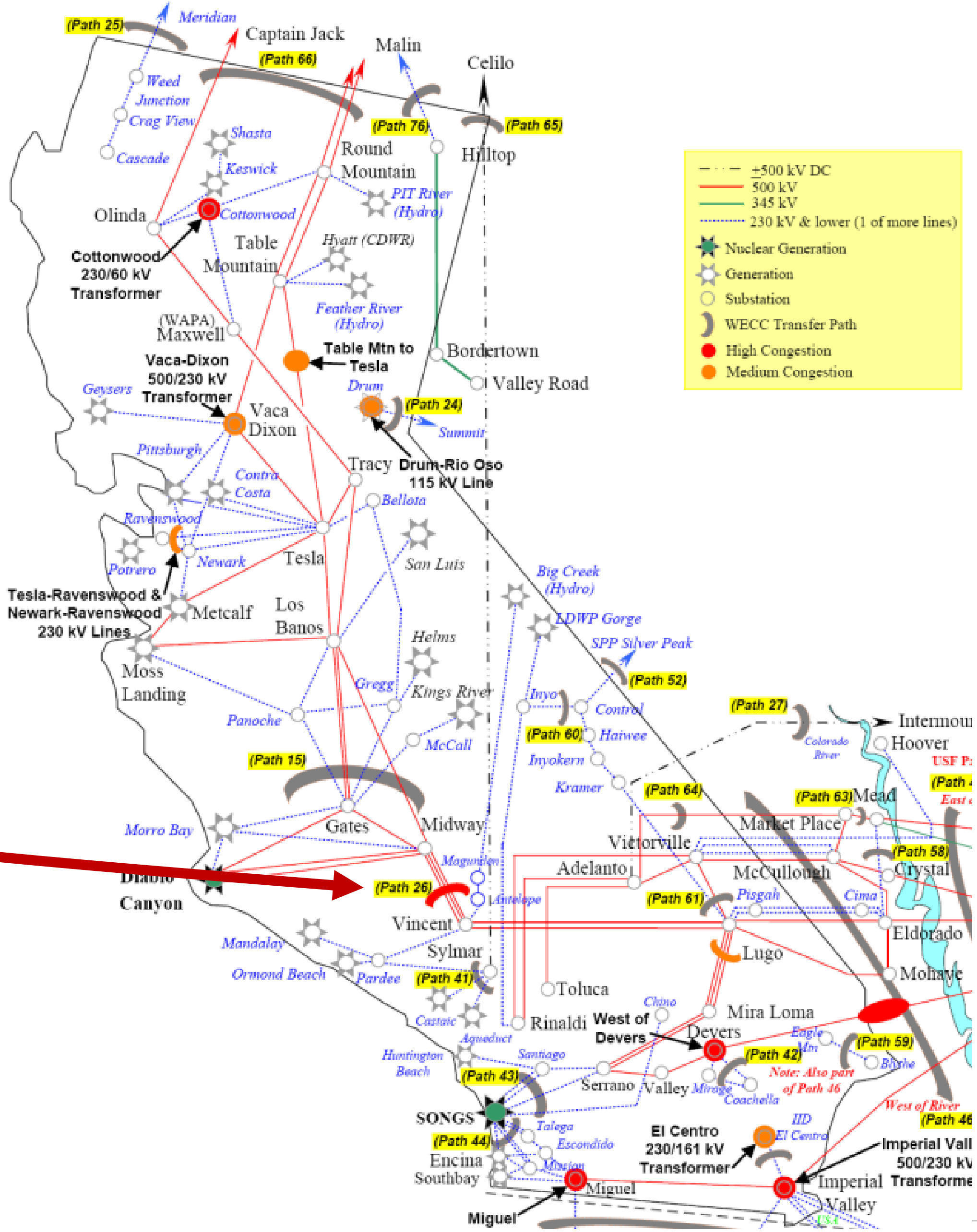
CART selects splitting variables and the logic of the tree. That is, CART selects PMU locations and the logic.

Decision Tree Diagram



PMU Placement: voting on the 3 Midway-Vincent lines

Three parallel 500 kV lines in path26



Heavy Winter training data.

- 4150 cases 133 measurements counting real and imag parts of currents. 43 voltage angles 40 complex current. Red vote Blue don't vote
  - Heavy Summer
- 11367 cases 113 measurements
- Voltage angles and real and imaginary parts of currents
- 15527 cases(rows) 246 possible measurements (columns)

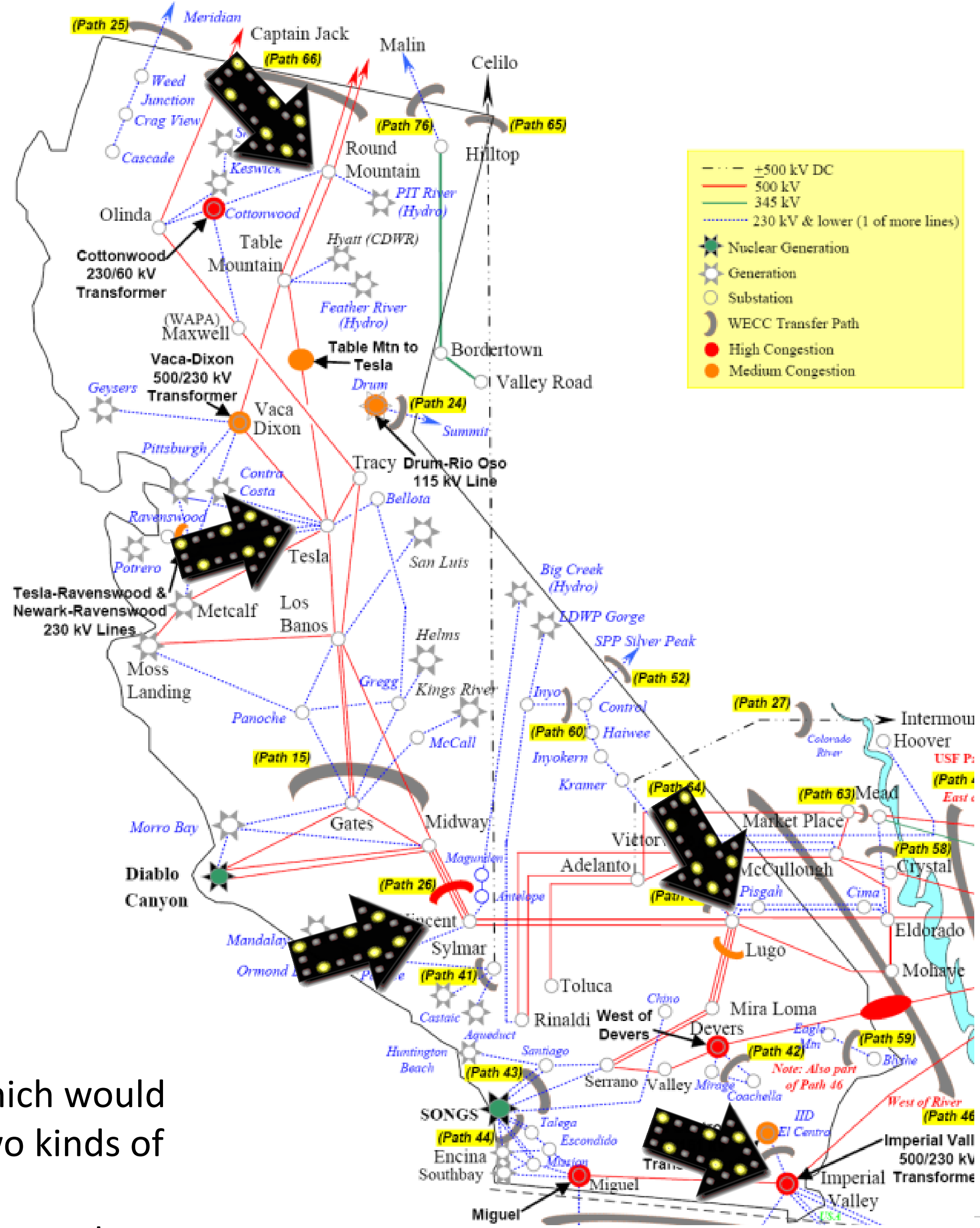
CART

- CART data is in an array, rows are events and outcome
- Columns are measurements. We use magnitude and angles for voltage and real and imaginary parts of currents. CART picks measurements to use for splitting.
- One column at a time the way we are doing it.
- If it picks the column it gets a real or imaginary part of the current to branch on. Creates a problem when you have to change the reference

PMU Placement: voting on the 3 Midway-Vincent lines

Line Current	
TESLA	LOSBANOS
LOSBANOS	MIDWAY
N.GILA	IMPRLVLY
TRACY	TESLA
LUGO	VINCENT
METCALF	MOSSLAND
LUGO	MOHAVE
MALIN	ROUND MT

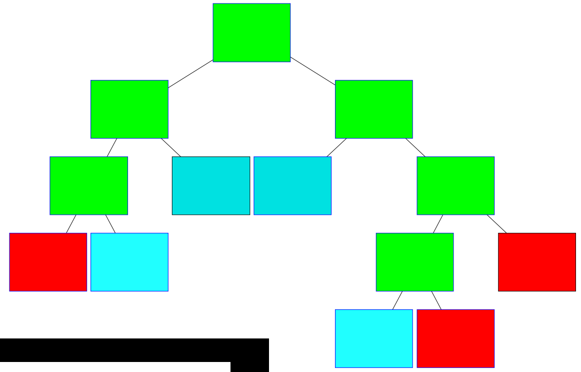
PMU
ROUND MT
TESLA
VINCENT
LUGO
IMPLVLY

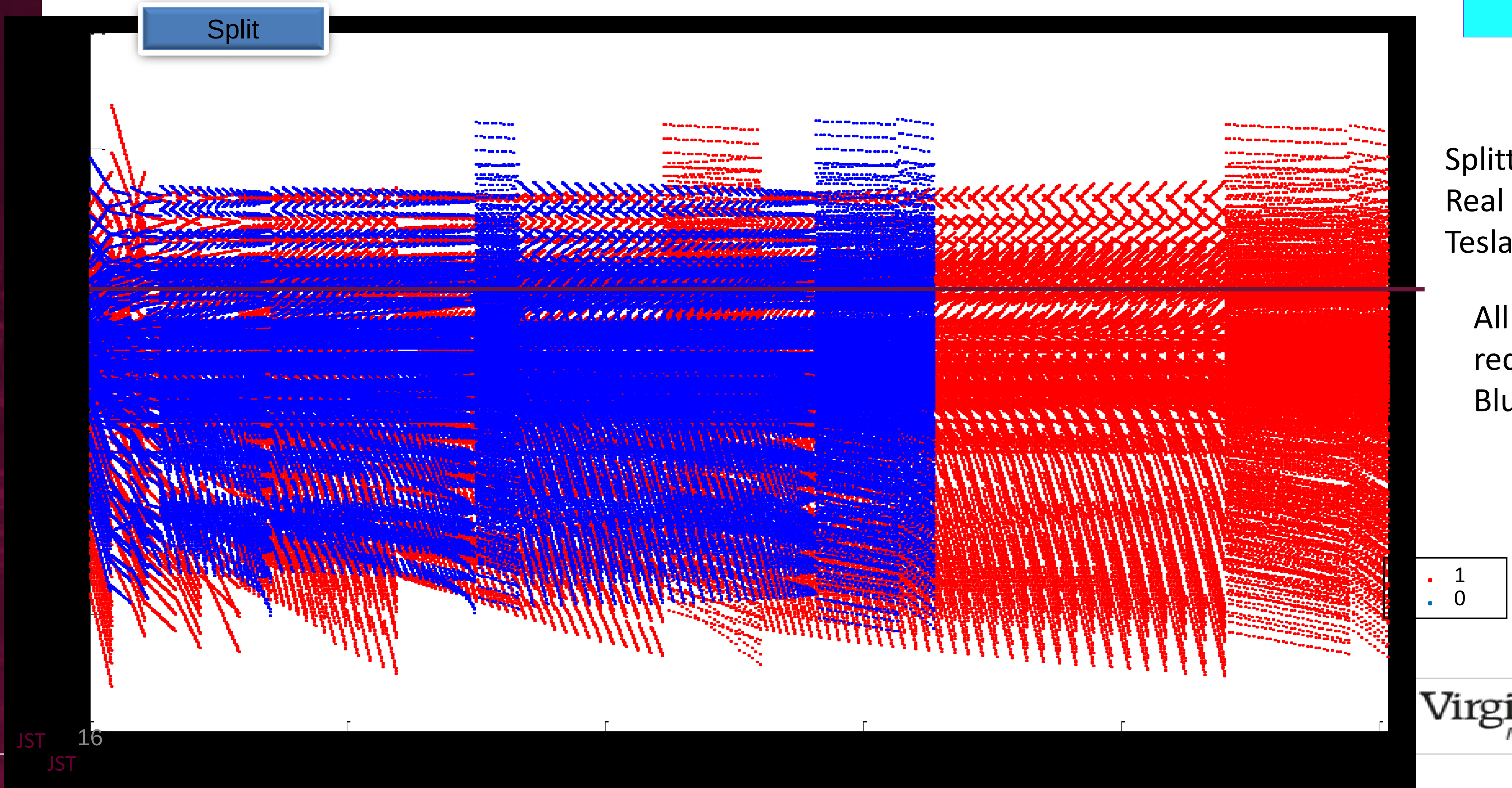


Less than 1% error rate in 15,000 cases half of which would have caused serious problems without voting. Two kinds of error: 1) fail to vote when you should have
2) Vote when you should not have. 1) is what we now do

VirginiaTech Invent the Future

What is CART doing?

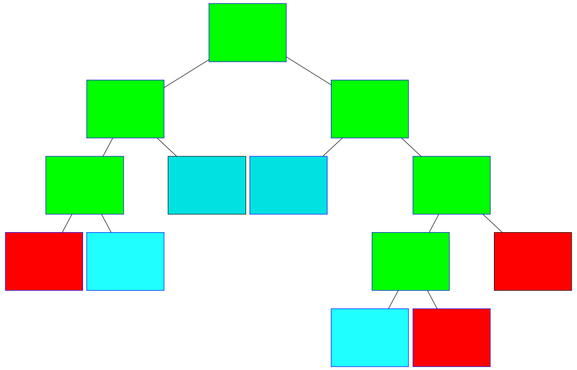




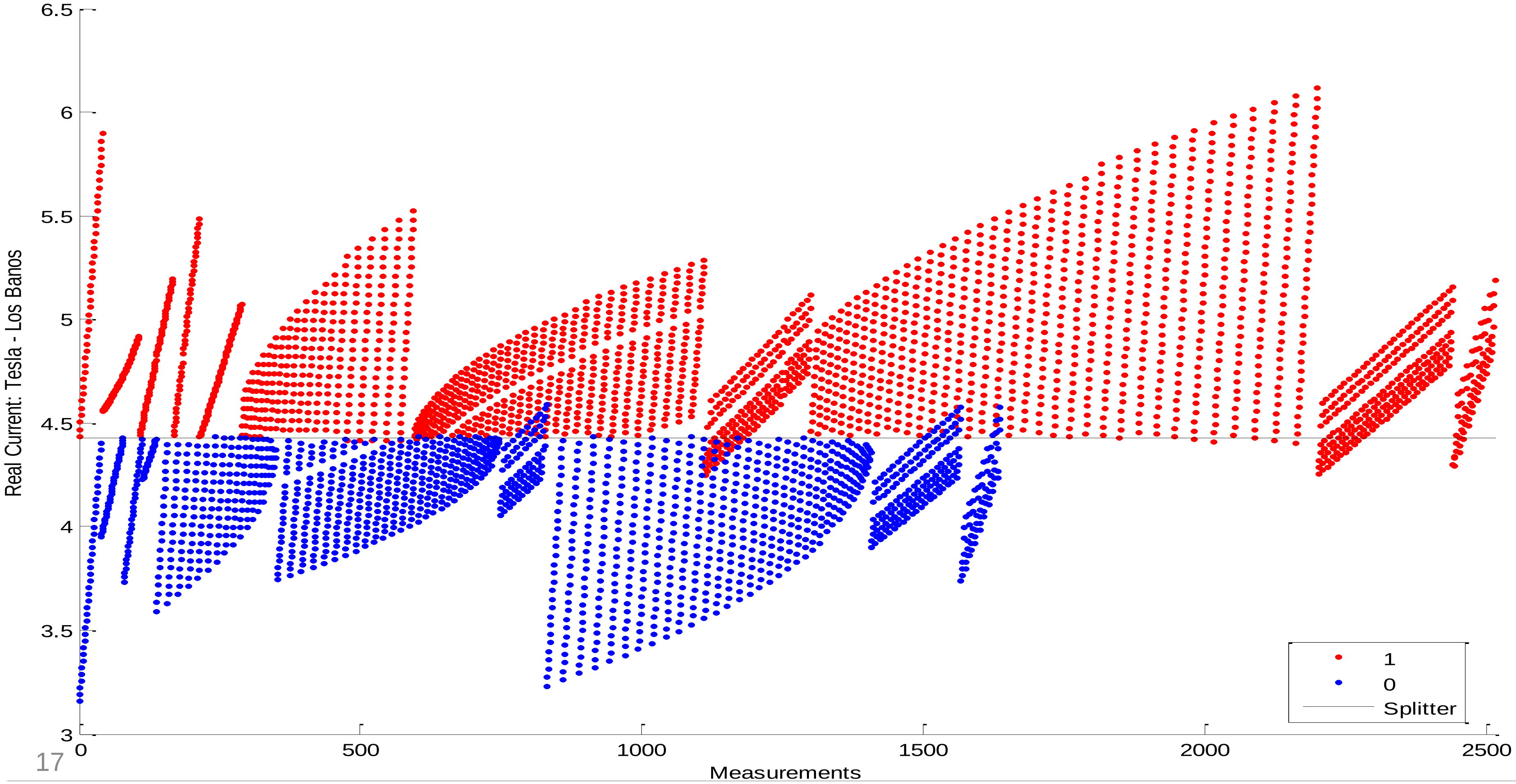
Splitter:
Real Current
Tesla – Los Banos

All the data
red vote
Blue don't vote

Node 1 Split:



- Predictor: Ir, Tesla – Los Banos

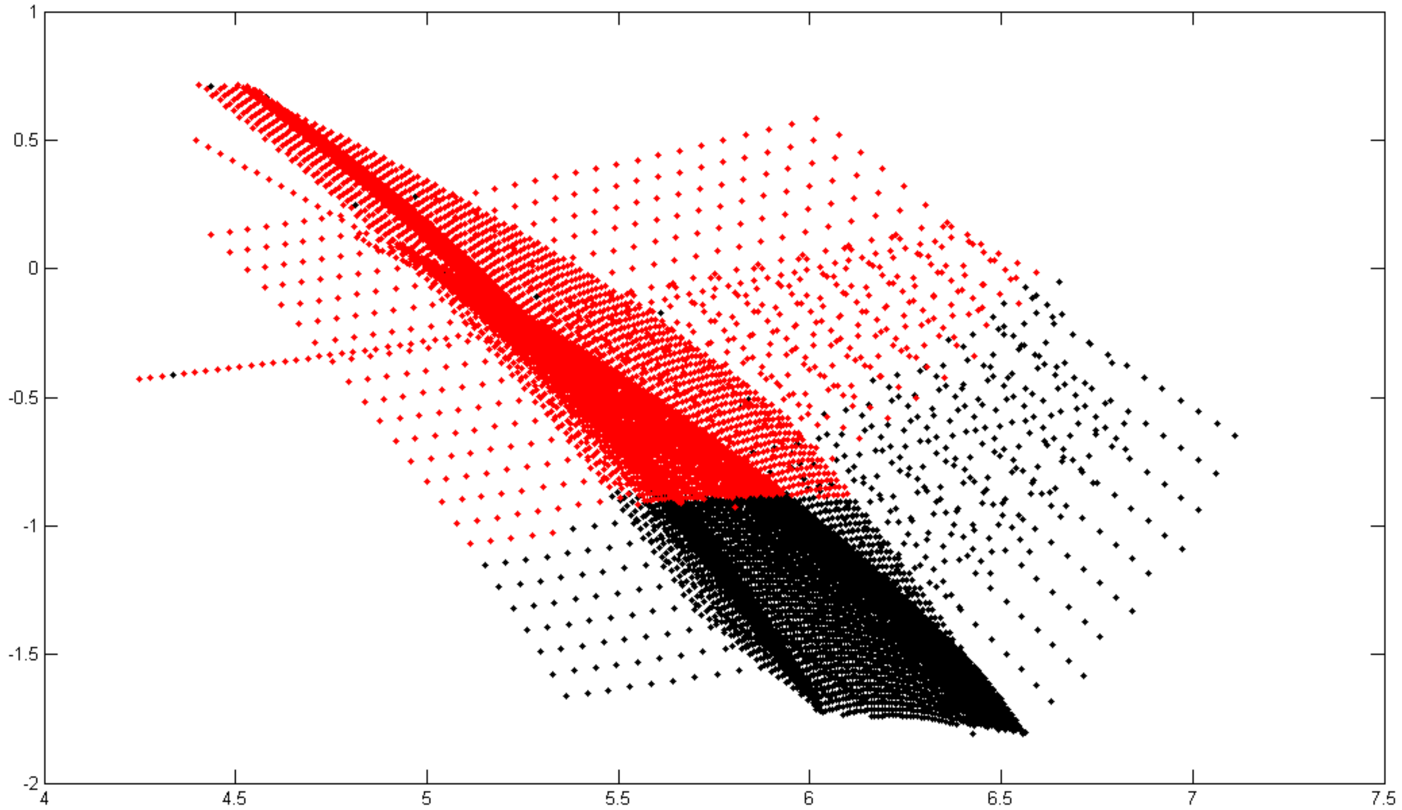


95% right with one split
The rest of the tree is for the remaining 5%

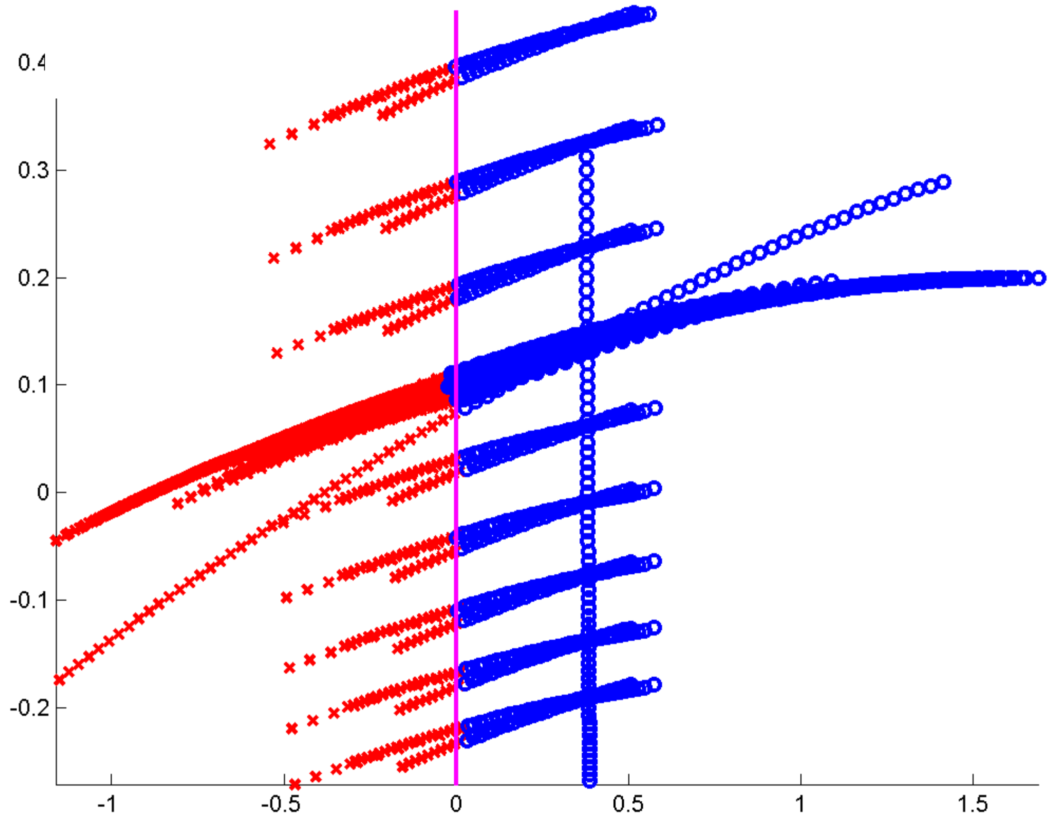
More about CART and the application

- Then PG&E said they did not want to waste a PMU at the reference bus Pittsburg. They wanted to make the reference bus a 500 kV bus.
- Could not find a 500 kV reference with the same 1% performance. Had to use different references in summer and winter
- I gave a couple of talks at Statistical and Applied Mathematical Sciences Institute (SAMSI) NSF, Duke, NCSU, UNC Consortium. They use CART to look for DNA cancer markers. My 15,000 ~ their 1,000,000

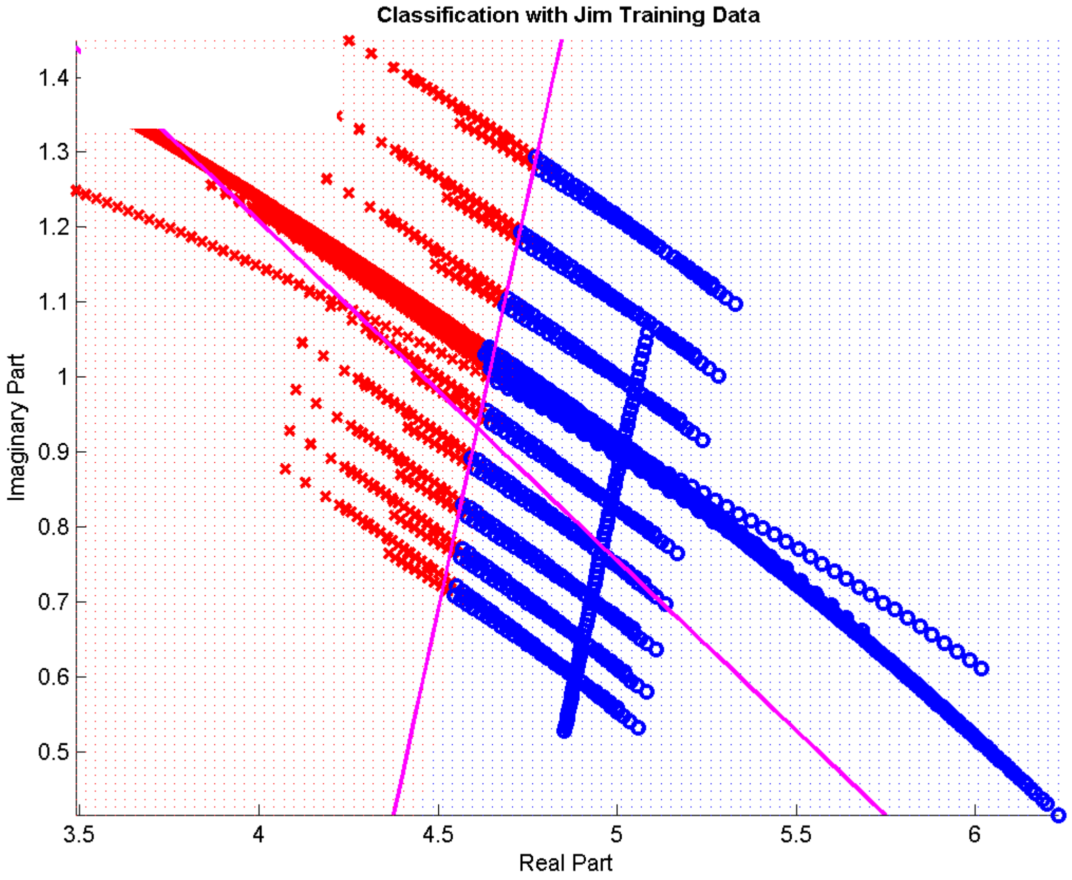
Real data



Would like to find data like this

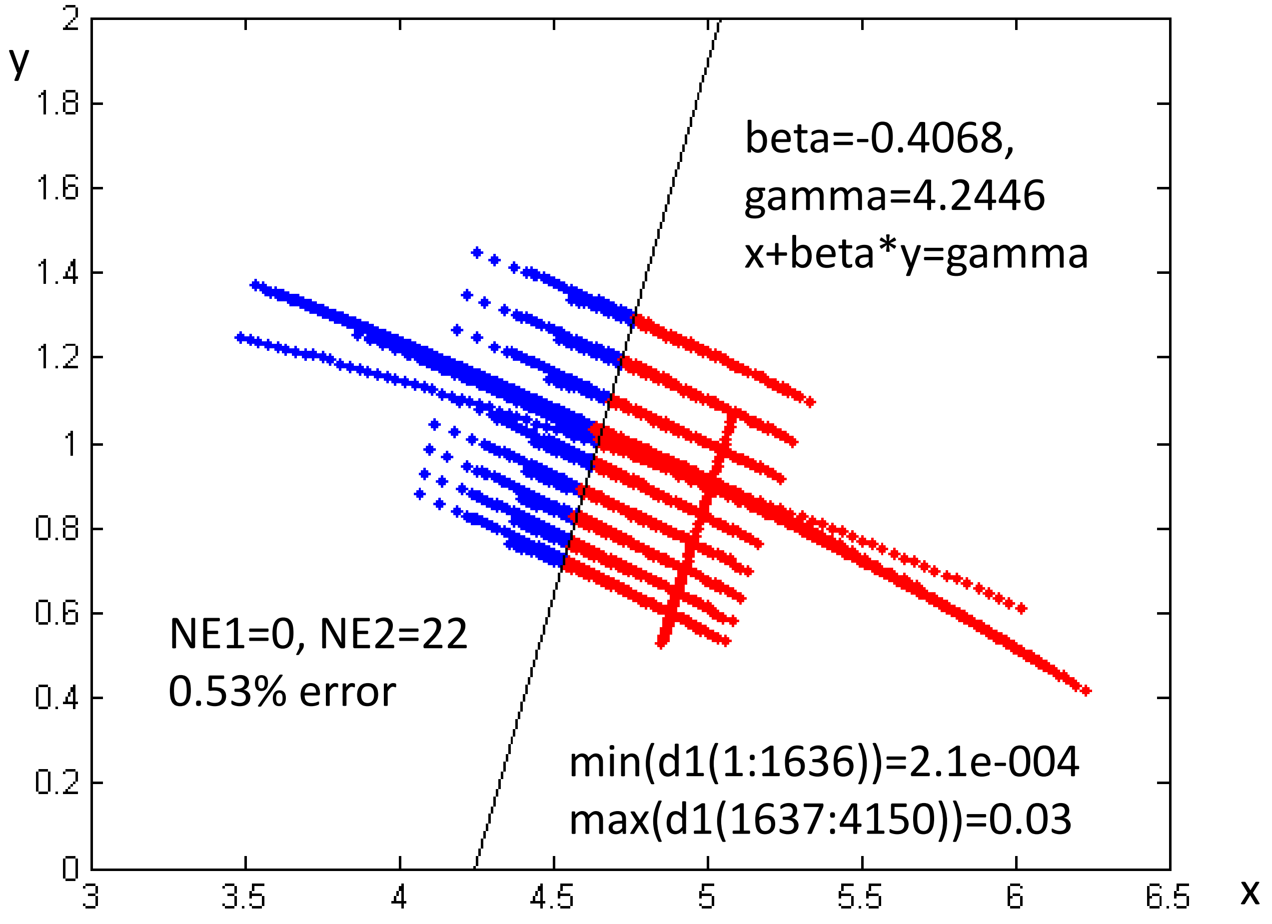


But this is what you are more likely to get

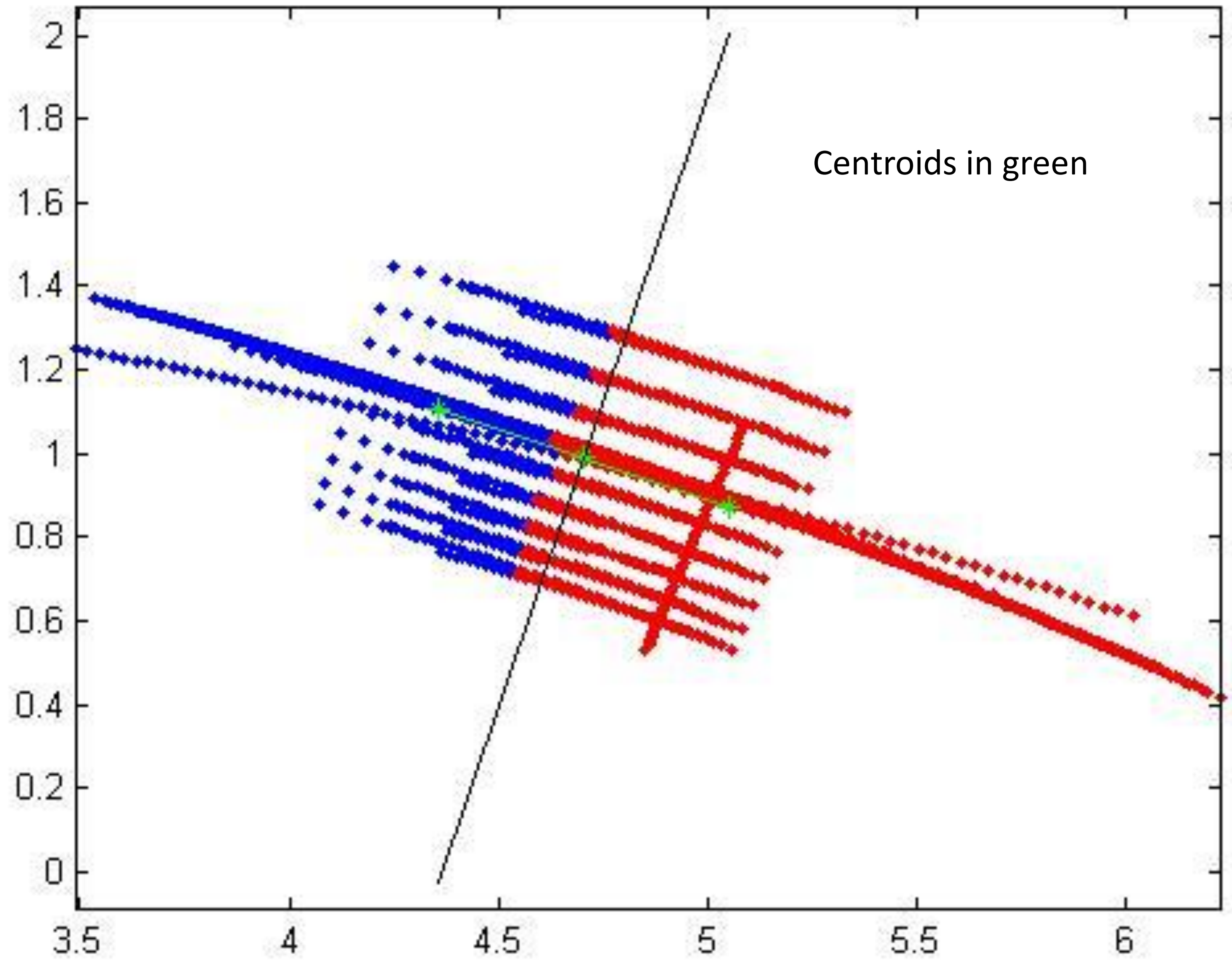


Solution

- Form the perpendicular bisector of the line joining the centroids.
- The centroid of the blue points is the average of the x point and the average of the y points taken over all 4150 points.
- Same for red points.
- Consider line joining the centroids. Bisect it and form a perpendicular

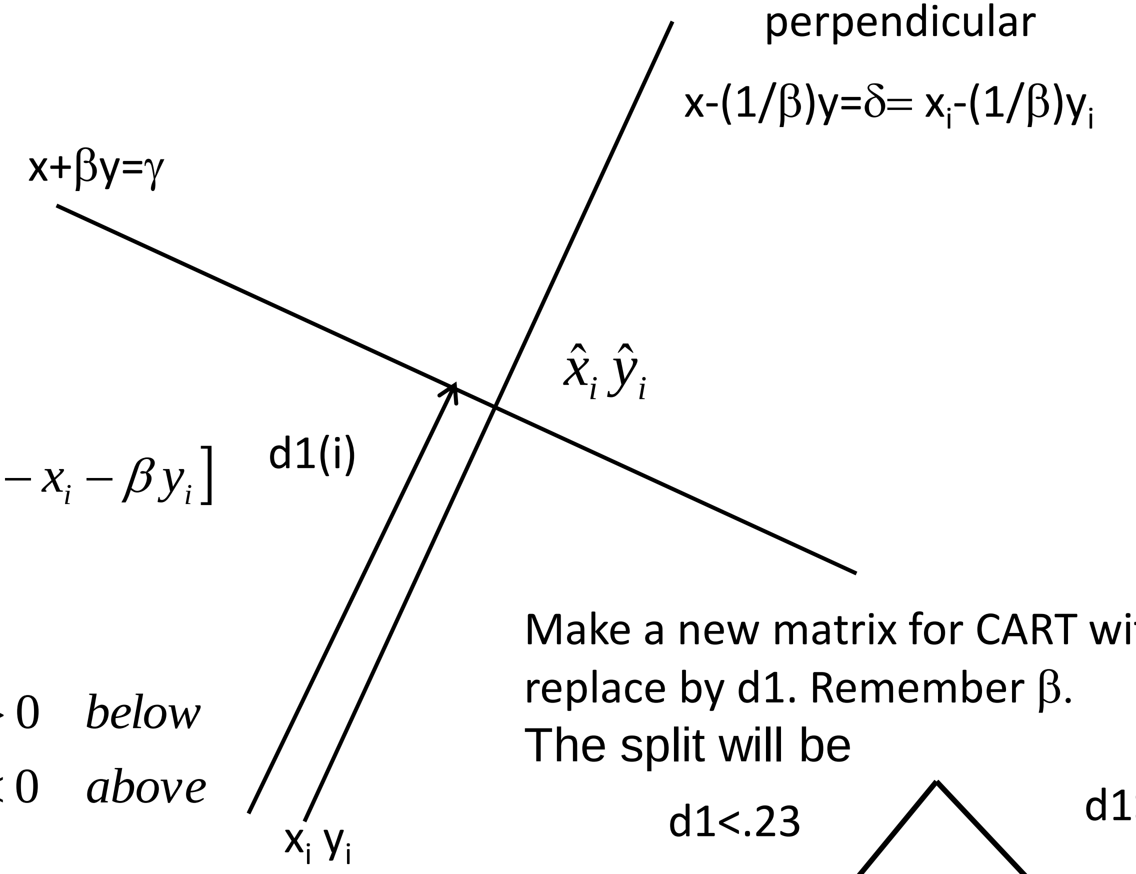


Heavy Winter line 1106 complex current per unit
Blue - don’t vote (1) 1636 points
Red - vote (0) 2514 points


Centroids in green
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
3.5
4
4.5
5
5.5
6

$x_{i,}\ y_i$ a data point

Dividing line



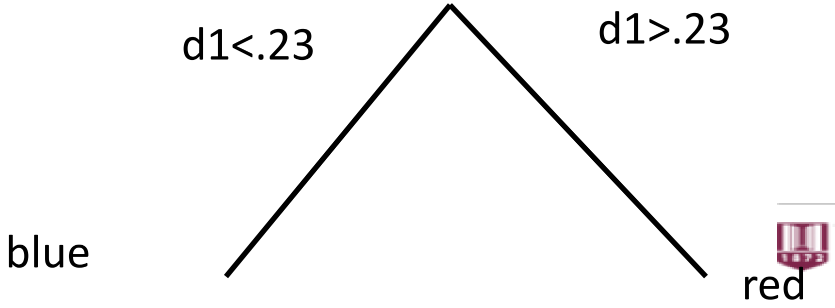
$$d1(i)=\frac{1}{(1+\beta^2)}\left[\gamma-x_i-\beta y_i\right]$$

$$sign(\beta d1(i))>0 \quad below$$
$$sign(\beta d1(i))<0 \quad above$$

$$d1(i)=\frac{1}{(1+\beta^2)}\left[\gamma-x_i-\beta y_i\right]>0.23$$

Make a new matrix for CART with x and y replace by d1. Remember β.

The split will be



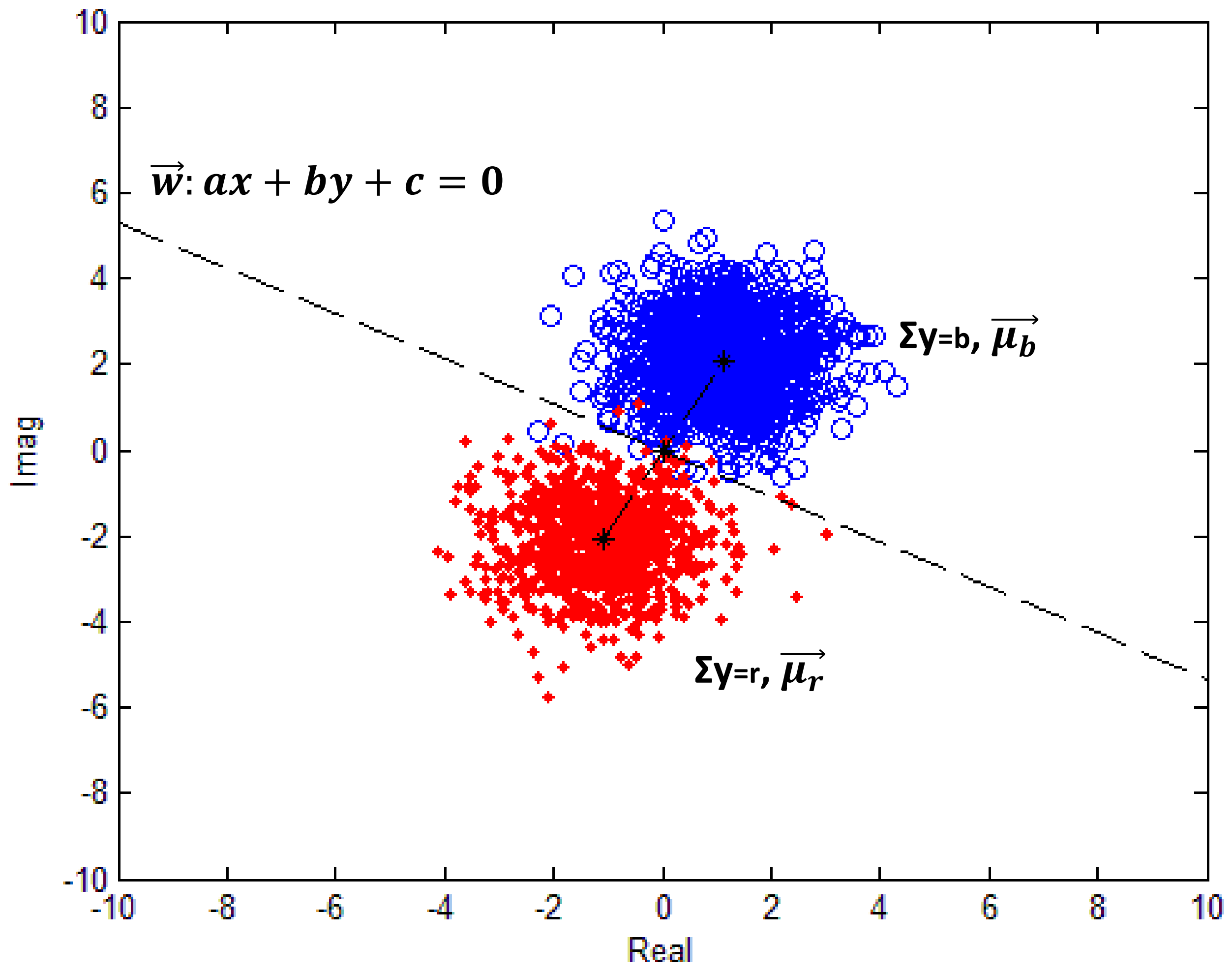
More dimensions

- A single PMU measures at least one voltage and one current (usually more). The minimum amount of data is 3 – a voltage angle and real and imag current. Could go up to 10 or 12.
- Trajectories in impedance. Even six point gives 12 numbers.
- Now need an idea from R. A. Fisher: Fisher’s Linear Discriminant Analysis (1936)[1] [2]
- Normalize the data with the experimental covariance matrix.
- That turns ellipsoids into spheres. Now the perpendicular bisector of the line joining the centriods is optimum

Classification Trees for Complex Data

- Traditional Decision Tree algorithms handle 1-D data making decisions based on a single attribute
- What we need:
  - Use the real as well as the imaginary components to make decisions
  - Extend the concept to make multi-class distinctions
- What we do:
  - Use Fisher's Linear Discriminant (FLD) to split complex data [1, 2]

Illustration of the Method



Mathematics Involved

- Equation of the separating hyper-plane is given by:

$$\left(\sum_{y=b} + \sum_{y=r}\right)^{-1} (\overrightarrow{\mu_b} - \overrightarrow{\mu_r})^T (\vec{x} - \vec{w}) = 0 \quad ----------------- -(6)$$

- Distance of the i^{th} point from the hyper-plane:

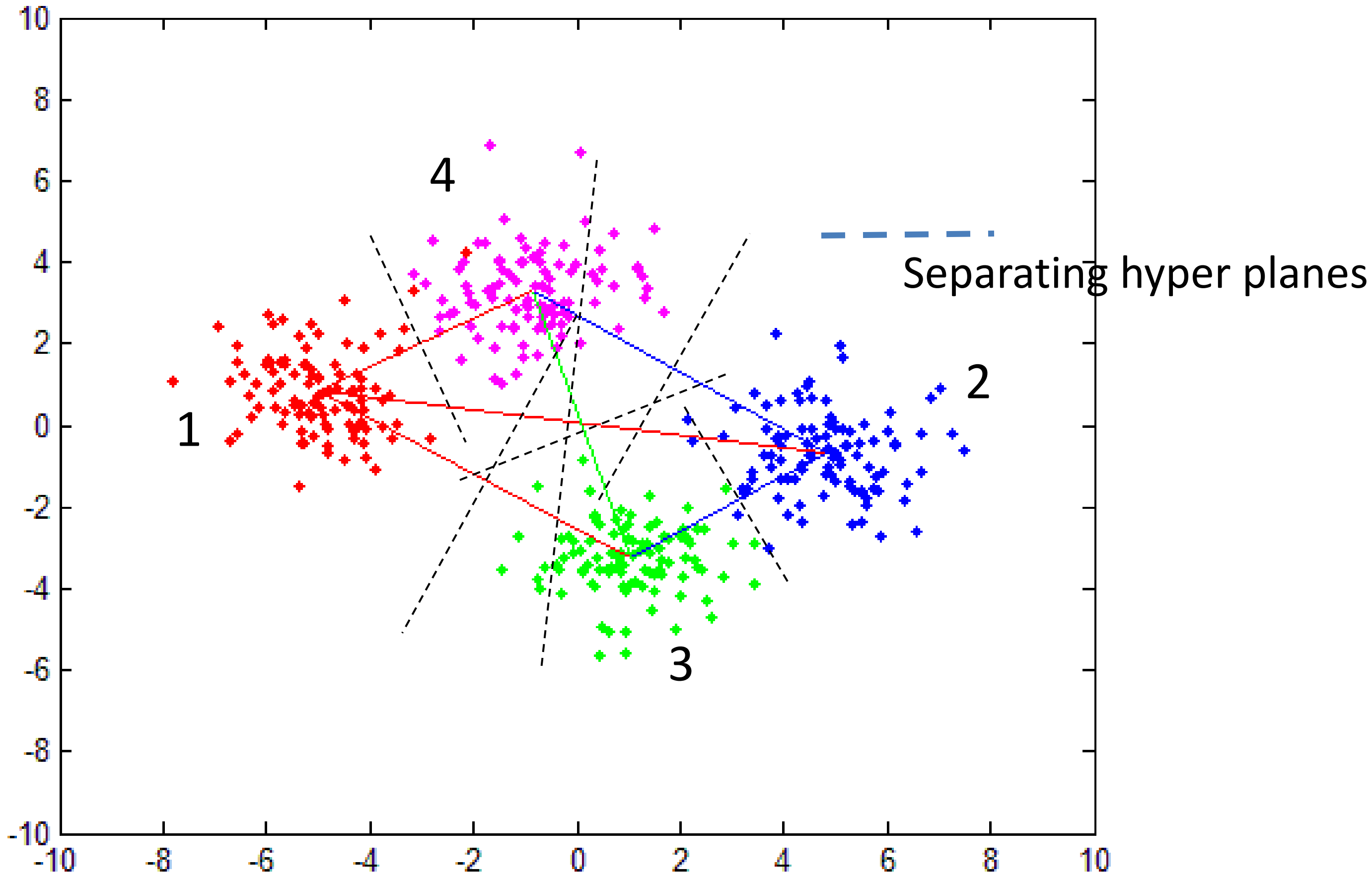
$$D(i) = \frac{ax_i + by_i + c}{\sqrt{a^2 + b^2}} \quad --------------------------------- -(7)$$

- New splitting variable, D:

$$\begin{array}{l} D(i) \leq 0: Red\ Dots \\ D(i) > 0: Blue\ Circles \end{array} \quad ------------------------------ -(8)$$

Extending to Higher Dimensions and Multi-Class Distinctions

4 classes of complex data 6 hyper planes, Splitting criteria distance to the hyper plane

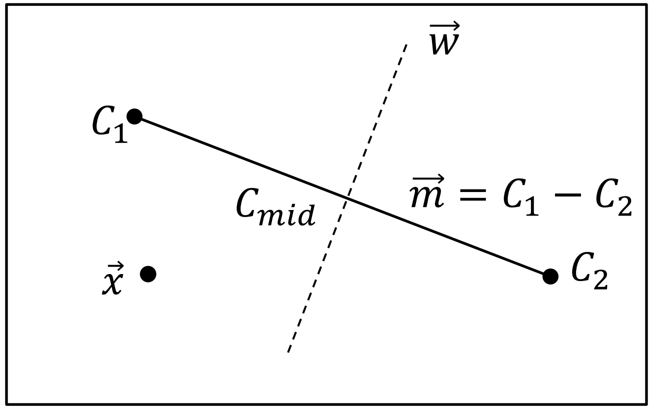


Number of Hyper-planes required $= \dfrac{\boldsymbol{n \times (n-1)}}{\mathbf{2}}$

Logic developed for handling high dimensional data

Any vector perpendicular to $\vec{m}$ is given by,

$$\frac{1}{2}(C_1 - C_2) + [\vec{w}\,][\alpha]$$



In order to find the optimum hyper-plane, we have to minimize: $\|\vec{x} - C_{mid} - [\vec{w}\,][\alpha]\|^2$

On solving, we get:

$$\boldsymbol{\alpha} = \left([\vec{\boldsymbol{w}}\,]^{\boldsymbol{T}}[\vec{\boldsymbol{w}}\,]\right)^{-\mathbf{1}}\left([\vec{\boldsymbol{w}}\,]^{\boldsymbol{T}}\right)(\vec{\boldsymbol{x}} - \boldsymbol{C_{mid}}) \quad ---------------------(9)$$

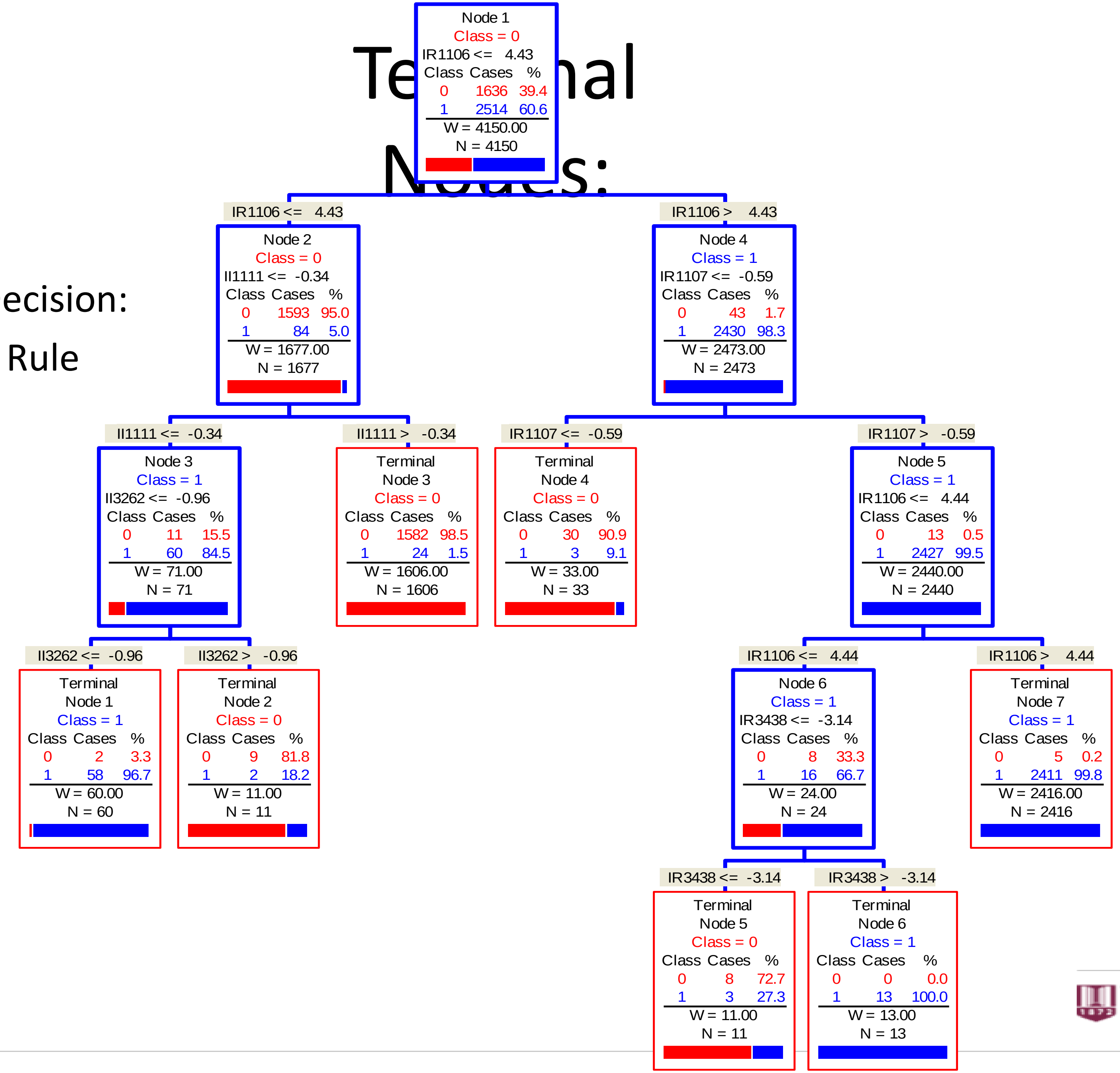
Substituting this value of α, we get the splitting variable in high dimensions as:

$$\vec{\boldsymbol{d}} = \left[\boldsymbol{I} - [\vec{\boldsymbol{w}}\,]\left([\vec{\boldsymbol{w}}\,]^{\boldsymbol{T}}[\vec{\boldsymbol{w}}\,]\right)^{-\mathbf{1}}\left([\vec{\boldsymbol{w}}\,]^{\boldsymbol{T}}\right)\right](\vec{\boldsymbol{x}} - \boldsymbol{C_{mid}}) \quad --------------(10)$$

Terminal Nodes:

Voting Decision:

Plurality Rule



Machine Learning in Power Systems: PMUs and Big Data

Used to construct algorithms that learn from data and are using that to make predictions or decisions. The data in power systems has come from simulation in the past and is on the threshold of using large amounts of archived data PMU. Machine learning is closely related to and often overlaps with a number of fields of computer science

- Computational Statistics
- Artificial Intelligence,
- Mathematical optimization
- Optical character recognition
- Data Mining
- Pattern Recognition
- Computer Vision

Our department at Virginia Tech now has a Machine Learning course which has a large number of power students enrolled.

VirginiaTech
Invent the Future

Classification and Regression

- Decision Trees CART
- Ensembles (Bagging, Boosting, Random Forests)
- Linear regression
- Naïve Bayes
- BART   Bayesian Regression Trees
- Neural networks
- Logistical regression
- Perceptron
- Support vector Machine SVM

Clustering

- BIRCH
- Hierarchical
- k-means
- EM
- DBSCAN
- Mean-shift

Anomaly Detection

- k-NN

Bagging decision trees, an early ensemble method, builds multiple decision trees by repeatedly resampling training data with replacement, and voting the trees for a consensus prediction

A Random Forest classifier uses a number of decision trees, in order to improve the classification rate.

Boosted Trees can be used for regression-type and classification-type problems

Outline of the rest

Hybrid Databases

Clustering

Other kinds of trees

SVM –Vladimir Vapnick Franklin Medal for Machine Learning

MapReduce

    Model Predictive control; Detrending Parallel

Anomaly detection

Base lining

Nearness to trouble

Regression

Need for adaptive Trees

The cloud

Hybrid Data Bases

The fact that major disturbances are rare makes it necessary to study some aspects of power system behavior with computer simulation. Particularly if the model validation of the previous section is accomplished there are significant advantages to be gained by combining real PMU data with simulation results to find a hybrid database that is a candidate for data mining. In the recent past most decision trees have been found using only data from simulations.

A slight variation is the work in [6] where simulations from a test system (10%) were merged with those of an actual operations planning model of the Hydro-Québec power grid (90%). While both are simulations they represent the issue of combining PMU data from different system models. The resulting database in [6] is skewed in the sense that vast majority of cases are stable. The solution in [6] is balanced by duplicating the unstable cases a number of times.

An alternate is to select simulation cases that will produce a balanced database as in [14]. That is, although the event is rare in the real world a data base that includes many extreme or rare events (balanced between stable or unstable) is appropriate for creating decision trees.

Clustering

There is a more limited history of data mining of archived PMU data. Fifteen months of PMU data of 54 angle differences has also been subjected to statistical analysis to detect abnormal power system behavior using software developed for NASA by PNNL [16-18].

Data is being mined to extract mode shapes, damping ratios, and frequencies facilitating establishing warnings/alarm thresholds for operator action. In addition all of the PMU measurements will be correlated with the system performance measures for normal operating conditions and its variations over a period of time and during various limiting conditions like thermal limits, proximity to voltage instability or voltage collapse, transient stability, etc.

This kind of analysis will facilitate establishing warnings/alarms thresholds for voltage phase angle measurements and determine site pairs of interest that are important in revealing systems stress (a natural CART function) and will recommend upper and lower limits for normal operation. It is argued [23] that the investment in improving monitoring of the high voltage transmission network represents the most cost-effective category of smart grid investment. But it is equally true that acceptance of the new technology is vital.

Other kinds of trees

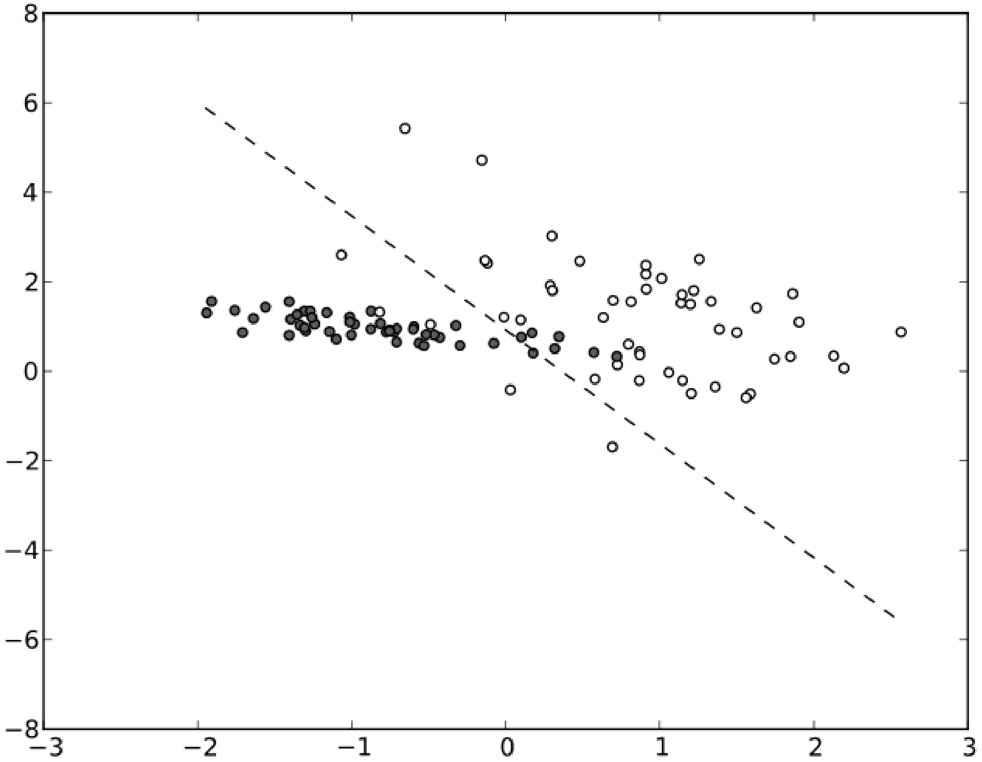
There are a number of alternate approaches to the creation of the decision tree. A number have been compared with a variety of criteria in [6].

They range from neural nets,
support vector machines,
random forests,
conventional decision trees
CART
Fuzzy decision trees, Fuzzy-IDT3 (Fuzzy Interactive Dichotomizer 3), and BART(Bayesian Additive Regression Trees) [22].
Certain elements of the creation of the database and the user's interaction with the tree are common however. For the various application envisioned the following summarizes the strategy to be employed.

SVM

- A support vector machine is a classifier that divides its input space into two regions, separated by a linear boundary. Here, it has learned to distinguish black and white circles

Detection of an outaged line is achieved using the variations of phase angles measured at the system buses where PMUs are located. Hence, protection from unexpected overloading in the network that may lead to system collapse can be achieved. Such detections are based upon an artificial intelligence technique which is the support Vector Machine (SVM) classification tool. [36]



MapReduce originally referred to the proprietary Google technology. It is now genericized and is part of Apache Hadoop an open source implementation. MapReduce is a programming model for processing and generating large data sets with a parallel, distributed algorithm on a cluster.

A MapReduce program is composed of a **Map()** procedure that performs filtering and sorting (such as sorting students by first name into queues, one queue for each name) and a **Reduce()** procedure that performs a summary operation (such as counting the number of students in each queue, yielding name frequencies). The "MapReduce System" (also called "infrastructure" or "framework") orchestrates the processing by marshalling the distributed servers, running the various tasks in parallel, managing all communications and data transfers between the various parts of the system, and providing for redundancy and fault tolerance.

Fast algorithms capable of processing massive volumes of data are now required in the field of power systems. An example is a parallel detrended fluctuation analysis (PDFA) approach for fast event detection on massive volumes of PMU data, taking advantage of a cluster computing platform[37]. The PDFA algorithm is evaluated using data from installed PMUs on the transmission system of Great Britain from the aspects of speedup, scalability, and accuracy.

'**Detrend**' In forecasting models, the process of removing the effects of accumulating data sets from a trend to show only the absolute changes in values and to allow potential cyclical patterns to be identified. This is done using regression and other statistical techniques.

VirginiaTech
Invent the Future

Anomaly Detection using MapReduce

The high rate of data samples reported by devices that support PMU functionality forces the use of non-traditional methods in order to attempt realtime anomaly detection. Two methods discussed are offline machine learning and a realtime sliding window procedure. In using machine learning techniques it is possible to assert a classifier algorithm, which to a certain degree of accuracy can flag incoming data for further operation when applied in realtime. The open source project Hadoop provides the storage architecture for large datasets (petabyte scale) as well as the MapReduce computational framework for distributed computing to produce these classifiers.[38]

Base Lining

NASPI's Planning and Implementation Task Team (PITT) made base lining of phase angle differences their highest priority. There has not been sufficient data available to date but [17] presents results using State Estimator data sampled every few minutes. Part of [17] is to identify data as atypical [18] rather than typical.

> If the PMU measurements can be correlated with system performance measures for normal operating conditions and its variations over a period of time and during various limiting conditions like thermal limits, proximity to voltage instability or voltage collapse, transient stability, etc. then decision trees can be formed important angle pairs identified.

The goal is to establish warnings/alarms thresholds for voltage phase angle measurements and determine site pairs of interest that are important in revealing systems stress (a natural CART function) and will recommend upper and lower limits for normal operation.

“Nearness to trouble”

The concept [17] of labeling different issues such as nearness to voltage instability or growing inter-area oscillations as both being either insecure or secure (I,S) can be used to classify each collection of data.

- By using the distance from separating hyper-planes between high dimensional collections of data (for example, a trajectory made up of complex voltage rather than a single measurement) as splitting variables in the trees [13] entire events can be labeled I or S and a tree constructed. Again simulation results from a validated system model will be used to make a database that is balanced.
- Note at this point the decision about which of the many techniques above to use is made. It may be that more field data may be used to connect the statistical concept of atypical behavior with power system concept of insecure. Are events that are insecure bordering on insecurity? Are there certain synchrophasor measurements that can serve as "bellwether" for identifying existing or imminent unit or system instabilities?

VirginiaTech
Invent the Future

Regression

<u>Estimation of line flow and voltages after an outage.</u>

An example of data mining in power system operations is described in [25] where an IDT3 tree is trained to predict the flows and bus voltages after outages in the Taiwan power system. There were no PMU measurements considered and all the cases were generated by simulation. The numerical values (P, Q, V) were quantized to 10 levels and the regression feature (the R in CART) was employed. The advantage was that in real time the tree could provide the new system state rapidly and accurately and contingency analysis was enhanced.

PMUs and The Cloud

There are obvious applications involving cloud computing envisioned [39-41], including the linear estimator [42]. The image of vast numbers of phasor measurements being delivered to the cloud and made available to the control center is certainly attractive. Attempts to place the open source linear estimator [43] in the cloud are currently underway . We are attempting to demonstrate the feasibility of running the open source linear estimator in the cloud.

The need for Machine learning that keeps learning

A tree or other answer obtained from simulation or archived data will be dated by power system growth and change.

Hydro Quebec has 60,000 cases that have been used for studies. It took 4 years to finish the adaptive –security-dependability project and install the implementation of the tree. It is likely no longer the best answer

The Hybrid data base where the information grows is one obvious requirement

We need to have a BART kind of response so that the tree evolves as the system changes. A Kalman Filter is inherently Bayesian –it continually adjusts the gain as data comes in. The CART tree has a structure that is determined by the algorithm and fixed.

Bagging decision trees, Random Forest classifiers, Boosted Trees and BART may hold the answer. Trees are only software and are not fixed but inputs are harder to change.

References

[1] R. Fisher, “The use of multiple measurements in taxonomic problems,” Annals of Eugenics, 7, pp. 179-188, 1936.

[2] Anamitra Pal, J. S. Thorp, Taufiquar Khan, and S. S. Young, “Classification Trees for Complex Synchrophasor Data,” , Electric Power Components and Systems, 41:1381–1396, 2013

[3] S. M. Rovnyak, S. E. Kretsinger, J. S. Thorp, and D. E. Brown, “Decision trees for real-time transient stability prediction,” IEEE Transactions on Power Systems, Vol. 9, no. 3, pp. 1417–1426, Aug. 1994.

[4] S. M. Rovnyak, C. W. Taylor, and J. S. Thorp, “Performance index and classifier approaches to real-time, discrete-event control,” Control Engineering Practice, vol. 5, no. 1, pp. 91–99, 1997.

[5] L.Wehenkel, T. Van Cutsem, and M. Ribbens-Pavella, “An artificial intelligence framework for on-line transient stability assessment of power systems,” IEEE Transactions on Power

[6] Kamwa I.; Samantaray, S.R.; Joos, G., “On the Accuracy versus Transparency Trade-Off of Data-Mining Models for Fast-Response PMU-Based Catastrophe Predictors”, IEEE Transactions on Smart Grid, Vol.: 1, Issue: 2, pp.: 144 – 158, 2010.

[7] Nuqui, R.F.; Phadke, A.G.; Schulz, R.P.; Bhatt, N., “Fast on-line voltage security monitoring using synchronized phasor measurements and decision trees”, IEEE PES Winter Meeting, 2001. Vol.: 3, pp.: 1347 – 1352, 2001.

[8] Ruisheng Diao; Kai Sun; Vittal, V.; O'Keefe, R.J.; Richardson, M.R.; Bhatt, N.; Stradford, D.; Sarawgi, S.K., “Decision Tree-Based Online Voltage Security Assessment Using PMU Measurements”, IEEE Transactions on Power systems Vol.: 24 , Issue: 2, pp.: 832 – 8, 2009.

[9] Sun, K., Likhate, S., Vittal, V, Mandal, S, “An Online Dynamic Security Assessment Scheme Using Phasor Measurements and Decision Trees”, Trans Power Systems, Vol. 22, No 4 2007 pp. 1935-1943

[10] Jan Ma; Makarov, Y.V.; Miller, C.H.; Nguyen, T.B.; “Use multi-dimensional ellipsoid to monitor dynamic behavior of power systems based on PMU measurement”, IEEE PES General Meeting, Conversion and Delivery of Electrical Energy in the 21st Century, pp.: 1 – 8, 2008

[11] Rui Sun; Zhongyu Wu; Centeno, V.A, “Power system islanding detection & identification using topology approach and decision tree” IEEE PES General Meeting, pp.: 1 – 6, 2011

[12] Kamwa, I.; Samantaray, S.R.; Joos, G., “Development of Rule-Based Classifiers for Rapid Stability Assessment of Wide-Area Post-Disturbance Records”, IEEE Transactions on Power systems, Vol.: 24, Issue: 1, pp.: 258 – 270, 2009.

[13] Zhiyong Li; Weilin Wu, ”Phasor Measurements-Aided Decision Trees for Power System Security Assessment”, ICIC '09. Second International Conference on Information and Computing Science, Vol.: 1, pp.: 358 – 361, 2009.

[14] Bernabeu, E.E.; Thorp, J.S.; Centeno, V., “Methodology for a Security/Dependability Adaptive Protection Scheme Based on Data Mining”, IEEE Transactions on Power Delivery, Vol: 27, Issue: 1, pp 104 – 111, 2012.

[15] Kamwa, I, Samantaray, S.R.,and Joos, G.,“Catastrophe predictors from ensemble decision-tree learning of wide-area severity indices,” *IEEE Trans. Smart Grid*, vol. 1, no. 2, pp. 144–158, Sep. 2010.

[16] Ferryman, TA, and Amidan, BG. 2010. “Statistical Analysis of Abnormal Electric Power Grid Behavior” 2010 Hawaii International Conference on System Sciences

VirginiaTech Invent the Future

[17] Ferryman, TA, and Amidan, BG,. “Investigation of Phase Angle Differences Using Statistical Analysis of Real World State Estimator Data,” 2012 Hawaii International Conference on System Sciences

[18] Amidan, B.G. and Ferryman, T.A., “ Atypical Event and Typical Pattern Detection within Complex Systems”, IEEEAC paper #1200, version 3 Dec 9, 2004

[22] Chipman,H.A.. George E.I., and McCuluch,R.E.,.“BART: Bayesian Additive Regression Trees”, *The Annals of Applied Statistics* 2010, Vol. 4, No. 1, 266–298

[23] Paul L Joskow, Creating a Smarter US Electricity Grid, CEEPR WP 2011-021, MIT Center For Energy and Environmental Policy Research, Oct 2011

[24] Liu, G., Quintere, J., Venkatatasubramanian, V.,”Oscillation Monitoring System Based on Wide Area Synchrophasors in Power Systems”, 2007 IREP Symposium Aug 19-24, 2007 Charleston SC

[25] Yang C-C, Hsu, Y-Y, “Estimation Of Line Flows and Bus Voltages Using Decision Trees” IEEE Transactions on Power Systems, Vol. 9, No. 3, August 1994,

[26] Thorp, J.S.; Ilic-Spong, M; Varghese, M, “Optimal secondary voltage-VAR control using pilot point information structure”, IEEE 1984 Conference on Decision and Control, Volume: 23 , Page(s): 462 – 466

[27] De La Ree, J.; Centeno, V.; Thorp, J.S.; Phadke, A.G., “Synchronized Phasor Measurement Applications in Power Systems” *IEEE Transactions on Smart Grid,* Volume: 1, Issue: 1 2010 , Page(s): 20 - 27

[28] Gao, Fenghua, Thorp, J.S., Gao, Shibin, “Dynamic State Prediction Based on Auto-Regressive Model using PMU Data” PECI 2012

[30] M. Maggioni, ***“What is...Data Mining?”,*** A.M.S. Notices, April 2012.

[31] W. K. Allard, G., Chen, and M. Maggioni, “Multi-scale geometric methods for data sets II: Geometric Multi-Resolution Analysis,” Applied and Computational Harmonic Analysis, Volume 32, pp. 435-462 (2012).

[32] Q. Wu, J. Guinney, M. Maggioni, S. Mukherjee, “*Learning gradients: predictive models that infer geometry and statistical dependence*, J.M.L.R., vol. 11 (August, 2010), pp. 2175-2198.

[33] N. G. Kingsbury, "Image processing with complex wavelets". *Phil. Trans. Royal Society London*. London (1999).

[34] L. Tao and A. Domijan, "On power quality indices and real time measurement," *Power Delivery, IEEE Transactions on* , vol.20, no.4, pp. 2552- 2562, Oct. 2005.

[35] Chen J, Thorp, J.S., Dobson, Ian, “Cascading dynamics and mitigation assessment in power system disturbances via a hidden failure model”, International Journal of Electrical Power & Energy Systems, Volume 27, Issue 4, May 2005, Pages 318–326

[36] Abdelaziz, A.Y.; Ain Shams Univ., Cairo, Egypt ; Mekhamer, S.F. ; Ezzat, M.; El-Saadany, E.F. Line outage detection using support Vector Machine (SVM) based on the Phasor Measurement Units (PMUs) technology, Power and Energy Society General Meeting, IEEE 22-26 July 2012

[37] Mukhtaj Khan, Phillip M. Ashton, Maozhen Li, Gareth A. Taylor, Ioana Pisica, and Junyong Liu, Parallel Detrended Fluctuation Analysis for Fast Event Detection on Massive PMU Data, IEEE TRANSACTIONS ON SMART GRID, VOL. 6, NO. 1, JANUARY 2015

[38]Paul Trachian, "Machine Learning and Windowed Subsecond Event Detection on PMU Data via Hadoop and the openPDC", Power and Energy Society General Meeting, 2010 IEEE

[39] Yufeng Xin ; Chakrabortty, A "A study on group communication in distributed wide-area measurement system networks in large power systems**",** Global Conference on Signal and Information Processing (GlobalSIP), 2013, Page(s): 543 - 546 IEEE Conference Publications

[40] Jian Yin; Sharma, P. ; Gorton, I. ; Akyoli, B. , "Large-Scale Data Challenges in Future Power Grids 2013 IEEE 7th International Symposium on Service Oriented System Engineering (SOSE), 2013 , Page(s): 324 – 328

[41] GIRI, J. Proactive Management of the Future Grid Power and Energy Technology Systems Journal, IEEE
Volume: PP , Issue: 99 Publication Year: 2015 , Page(s): 1

[42] Thoshitha Gamage, David Anderson, David Bakken, Kenneth Birman, Anjan Bose, Carl Hauser, Ketan Maheshwri, and Robbert van Renesse. Mission-Critical Cloud Computing for Next-Generation Power Applications. In *Smart Grids: Clouds, Communications, Open Source, and Automation*. Editors David Bakken and Kris Iniewski. CRC Press. ISBN 9781482206111. 2014.

[43] K. D. Jones, J. S. Thorp, and R. M. Gardner, "Three-phase linear state estimation using phasor measurements," in *Proc. IEEE Power Eng. Soc. Gen. Meeting*, Vancouver, BC, Canada, pp. 1-5, 21-25 Jul. 2013.

VirginiaTech
Invent the Future