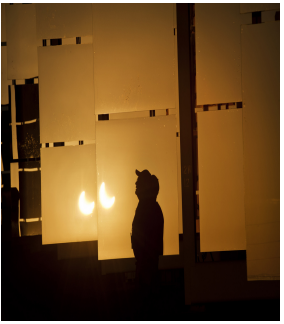
5/10/17



PV Monitoring and Modeling

a machine learning perspective

C. Birk Jones, Ph.D
(cbjones@sandia.gov)

Sandia National Laboratories

Exceptional service in the national interest

U.S. DEPARTMENT OF ENERGY

SunShot U.S. Department of Energy

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy?s National Nuclear Security Administration under contract DE-NA0003525. SAND NO. 2017-624055

Overview

What is Machine Learning?

What?

Machine Learning is the science of programming a computer so that it can learn from data

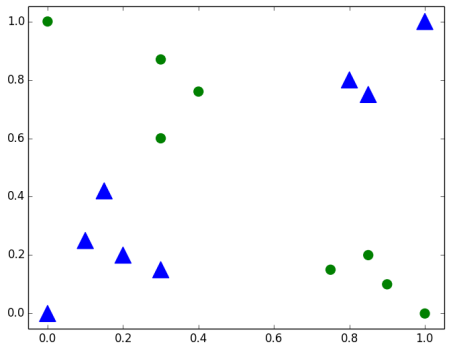
Two main types:

1. Supervised learning:
   Inputs w/ desired outputs
2. Unsupervised learning:
   Inputs w/out desired outputs

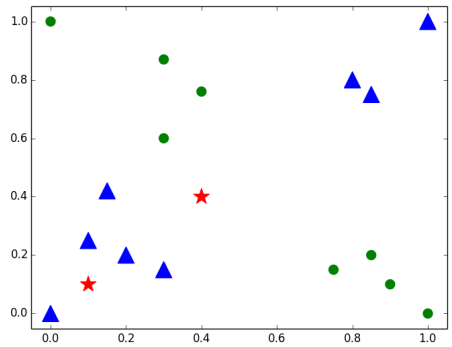
Supervised Learning Example

Classification using Support Vector Machine

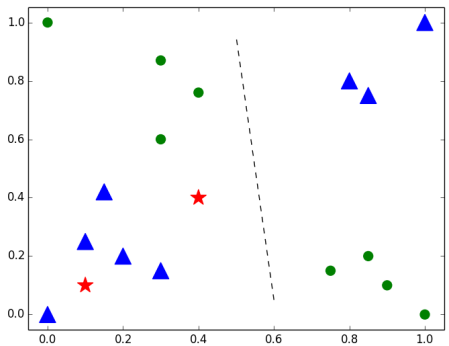
XOR data



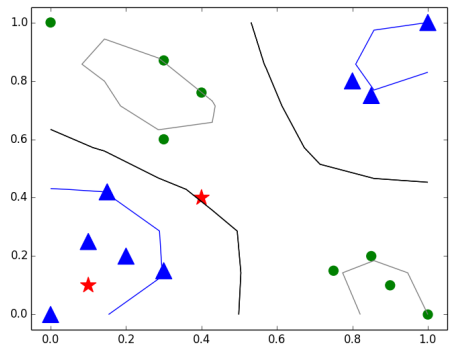
Classify new data



Not linearly seperable



Non-linear classification



Why use Machine Learning?

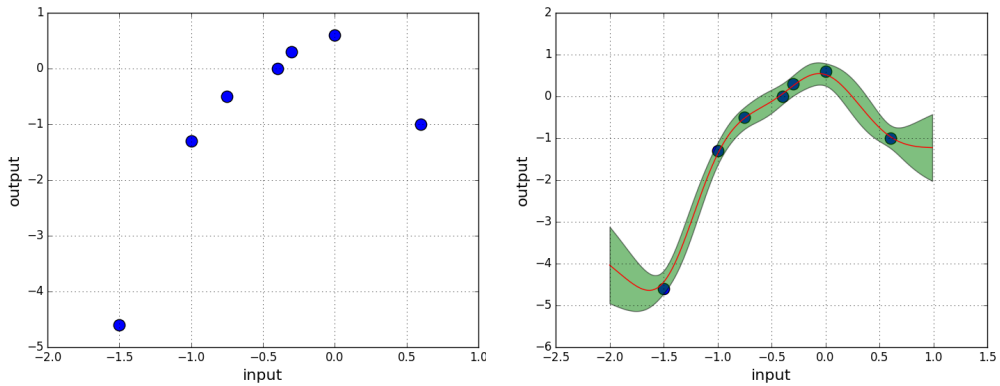
Why?

1. Regression
2. Classification
3. Density estimation
4. Others

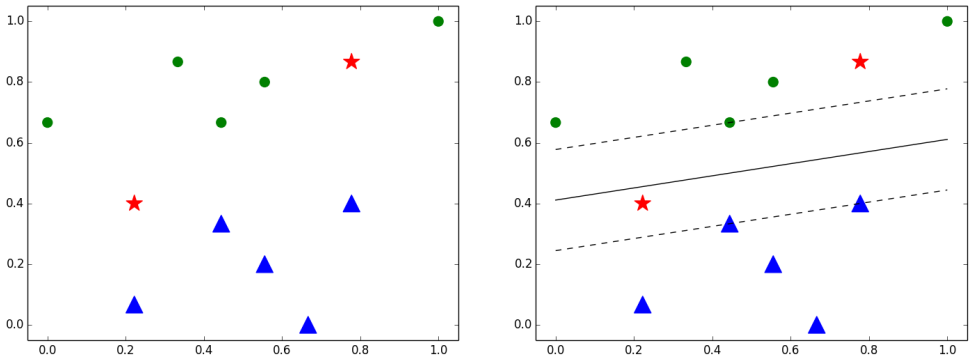
Approaches:

1. Artificial Neural Networks
2. Deep Learning
3. Support Vector Machines
4. Clustering
5. Many others

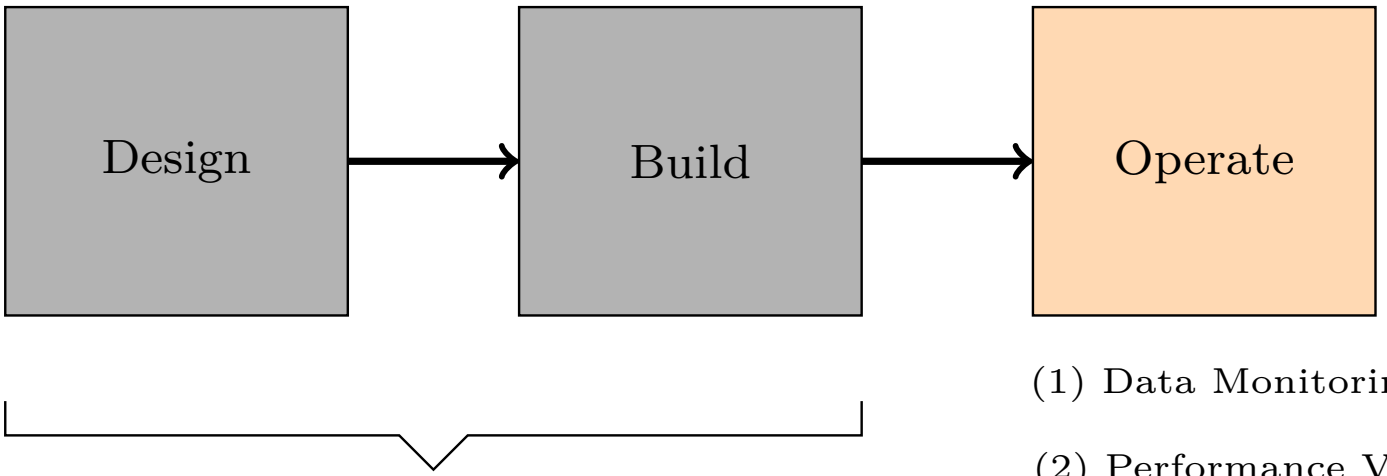
Regression (Gaussian Process)



Classification (Support Vector Machine)



How can Machine Learning Provide Value?



(1) Data Monitoring?

(2) Performance Validation?

(3) Identify Faults?

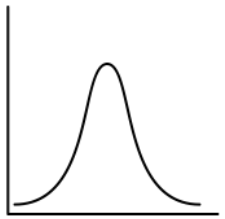
Machine Learning Applications?

Data Monitoring Quality:

Rule-based

```
if True then
    Report
else
    No Report
end if
```

Outlier Detection



Validate Performance:

Extensive Data

Inputs = {E,T_{module}}

Outputs = {Power}

Limited Data

InputsA = {P_{nearby} }

InputsB = {$E_{forecast}$}

Outputs = {Power}

Fault Analysis:

Fault Detection

Anomaly/Novelty

Detection

Fault Diagnostics

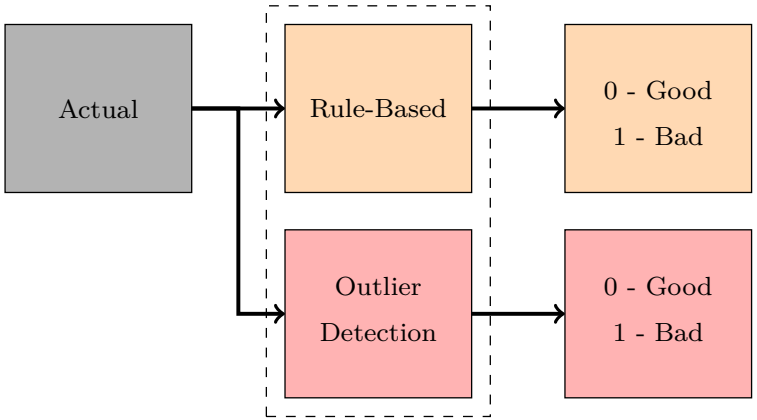
Multi-Class Classification

Data Monitoring: Rule-Based versus Outlier Detection

Problem Statement

Machine learning can use polluted data that contains incorrect or corrupt data records to identify outliers by assuming a distribution.

Process



Results

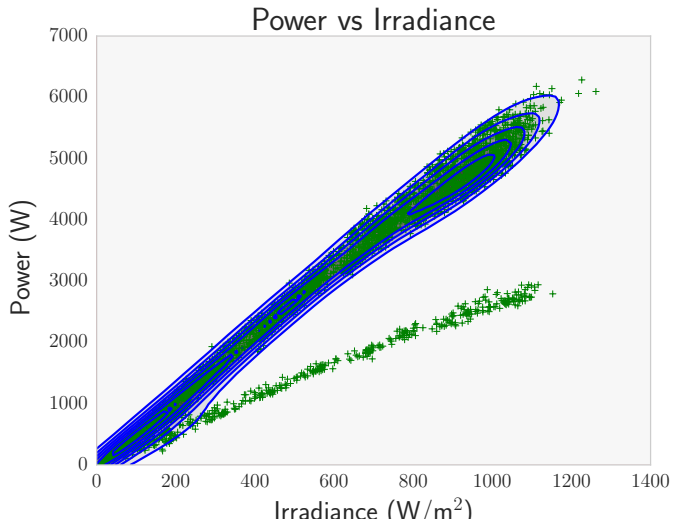
Rule-Based

Rules were not violated

if Power $>$ P_{mpp} x $Number_{mod.}$ x 1.2 $=$ 7776
then
    Alarm
else if Power $<$ 0 then
    Alarm
end if

Kernel Density Function

Outliers were detected

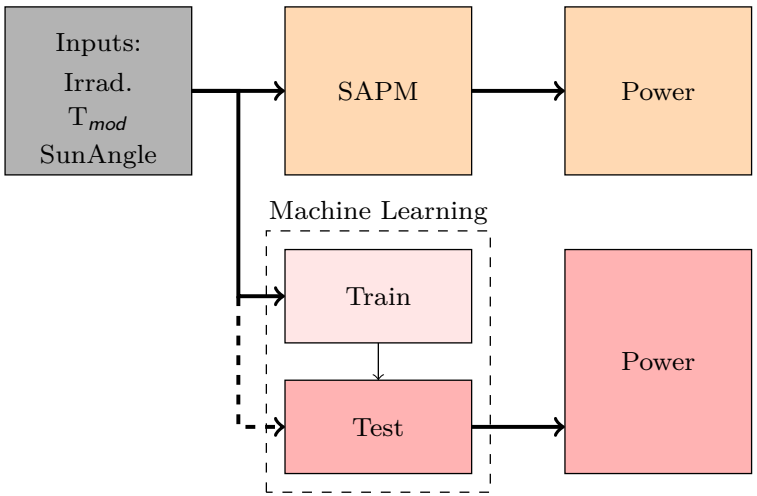


Validate Performance: Extensive Data

Problem Statement

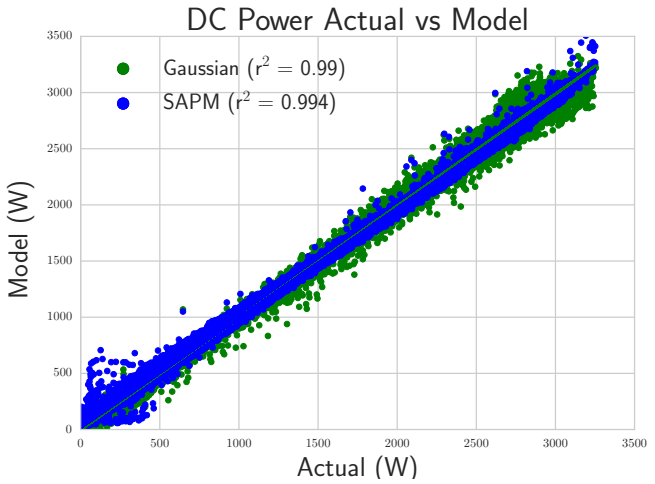
Machine learning can model existing PV systems using collected weather and performance data.

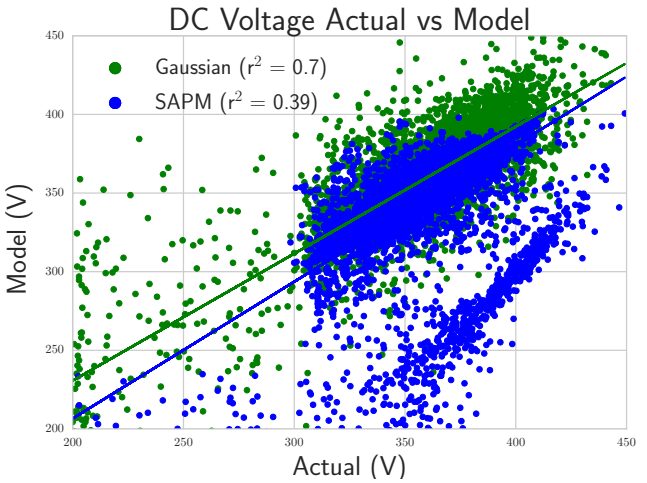
Process



Train: 01/16 to 12/16 -> Test: 01/17 to 04/17

Results



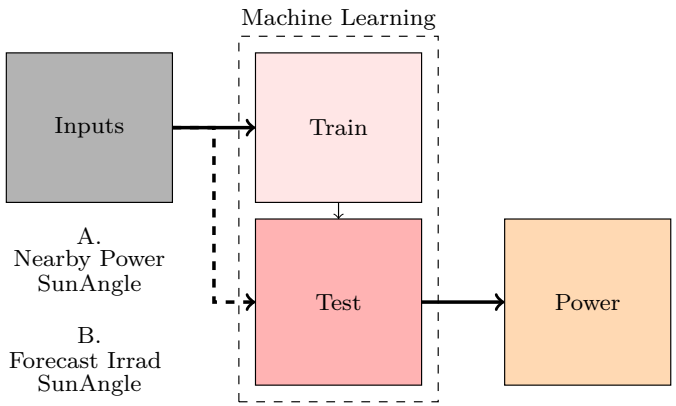


Validate Performance: Limited Data

Problem Statement

Machine learning can model existing PV systems where power is the only monitored value. The algorithm can associate PV power with nearby system outputs and forecasted irradiance.

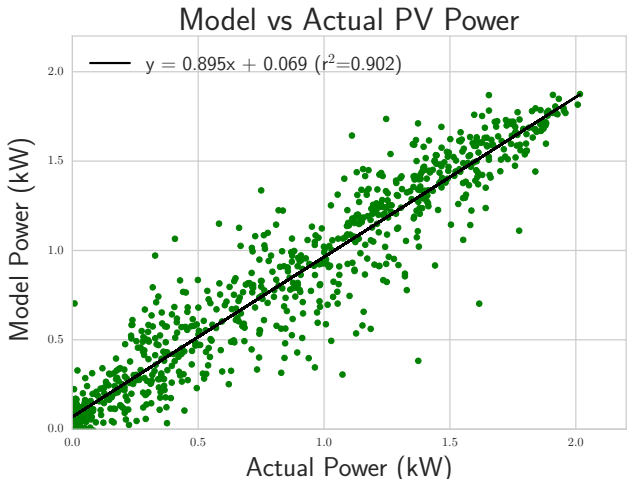
Process



Results

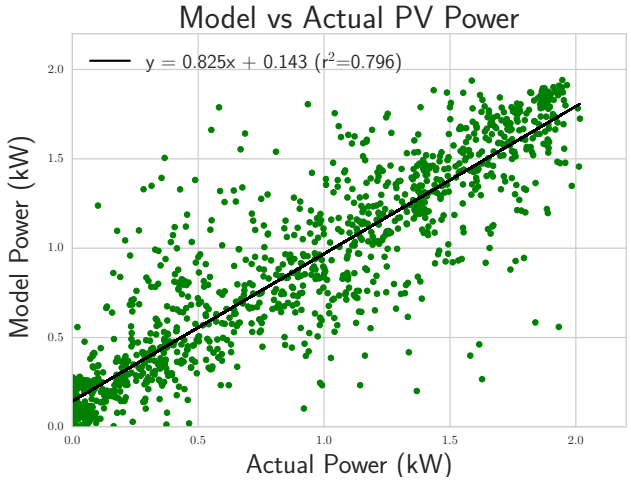
Inputs A:

1. Nearby Power (17km)
2. Sun Angle



Inputs B:

1. Forecast Irrad. (NOAA)
2. Sun Angle



FDD: Novelty Detection (I-V Data)

Problem Statement

Machine learning can be used to perform binary classification of I-V curve data.

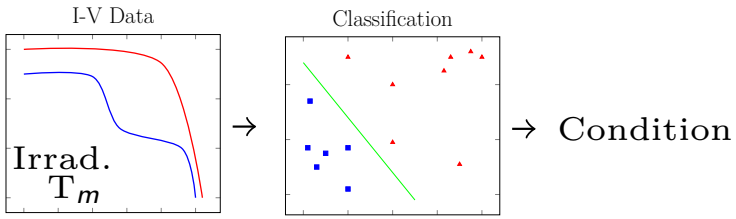
Process

Train & Test

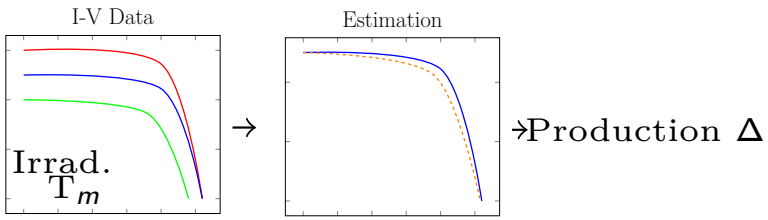
Training: 86 normal and 21 abnormal curves

Testing: 242 normal and 82 abnormal curves

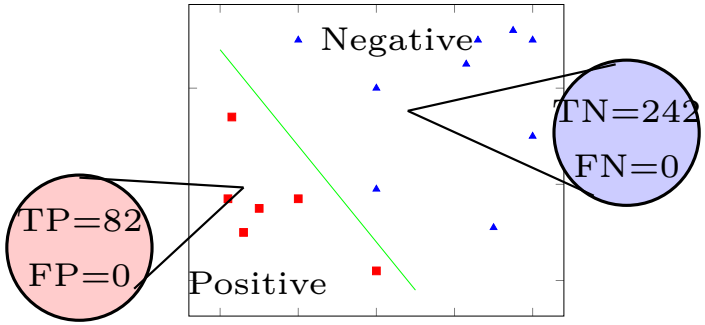
Classification



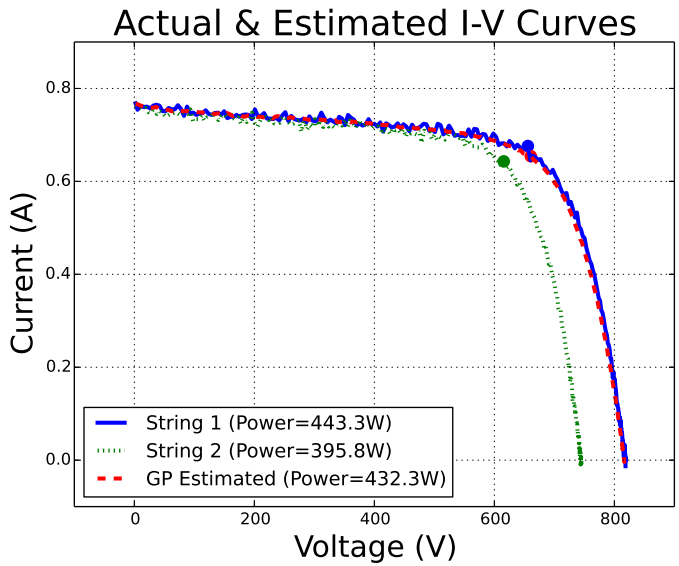
Estimation



Classification



Estimate Loss

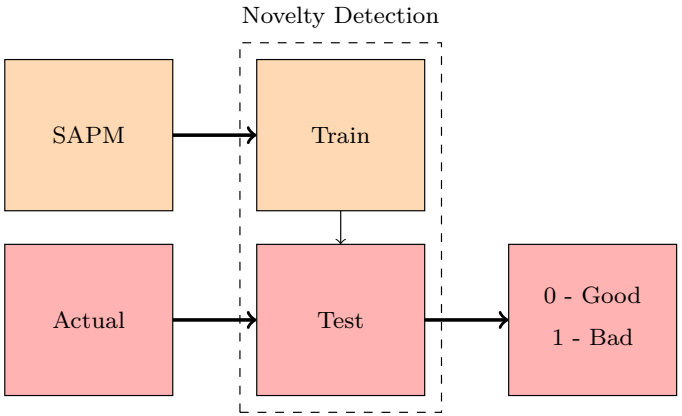


FDD: Novelty Detection (Max Power Point Data)

Problem Statement

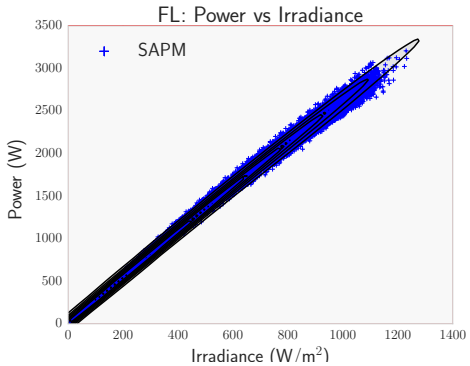
Machine learning can be used to identify anomalies automatically by training on "clean" data and testing on new, possibly polluted, observations.

Process

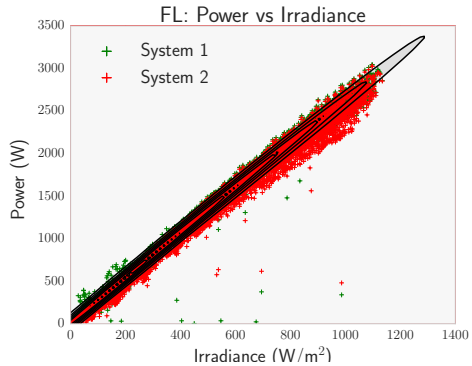


Train & Test Results

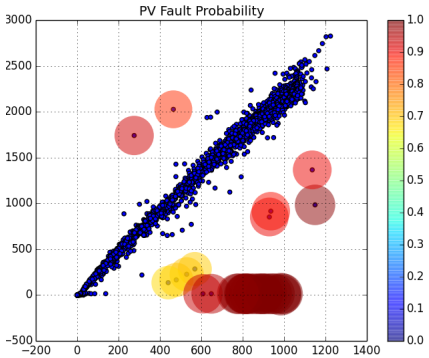
Train



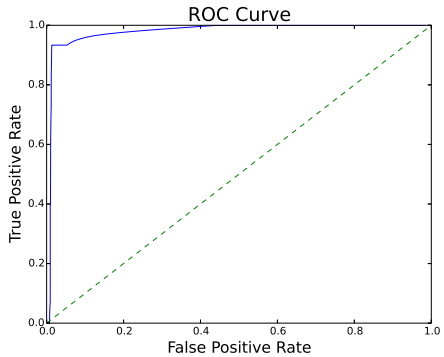
Test



Estimate Probabilities



Evaluate Accuracy



FDD: Classification using LAPART

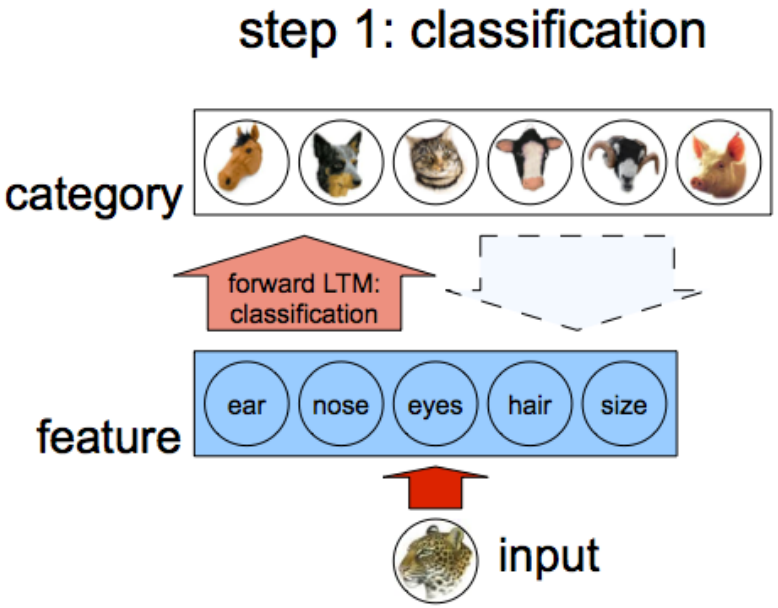
What?

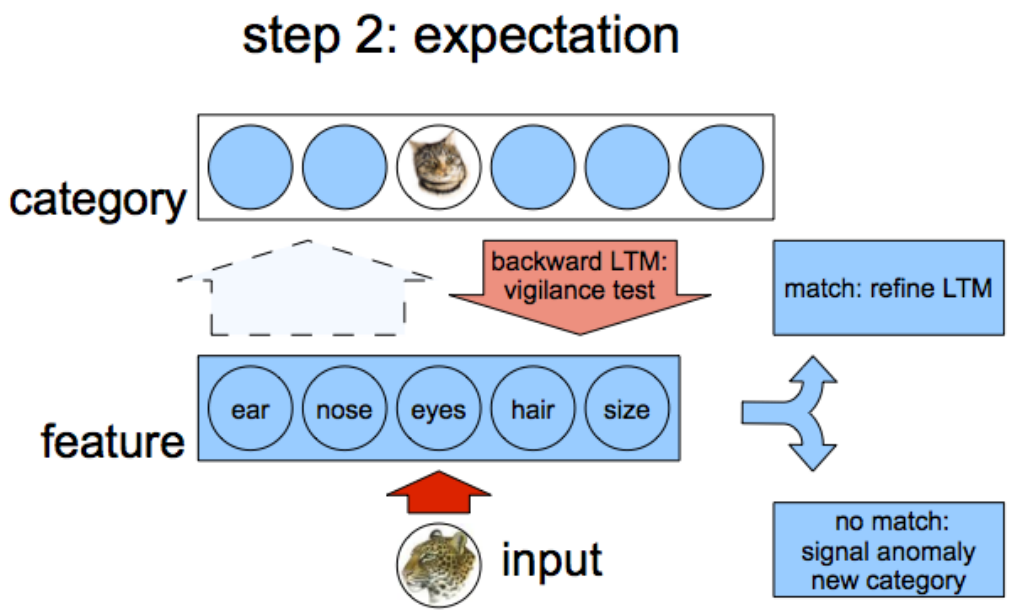
1. Unsupervised
2. Supervised

Why?

1. Classification
2. Regression

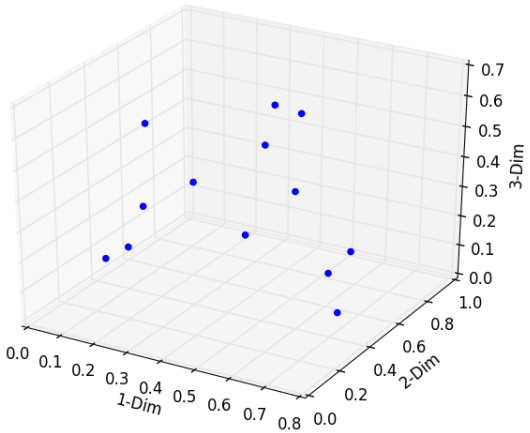
How?

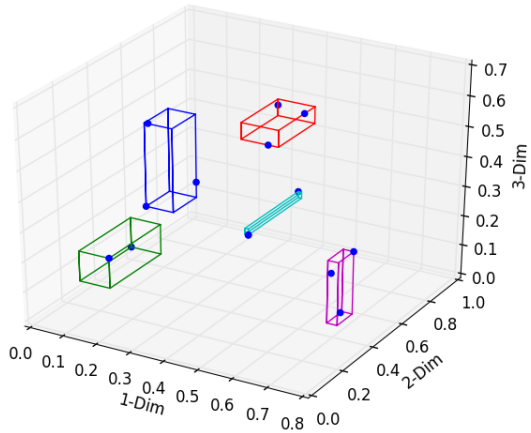




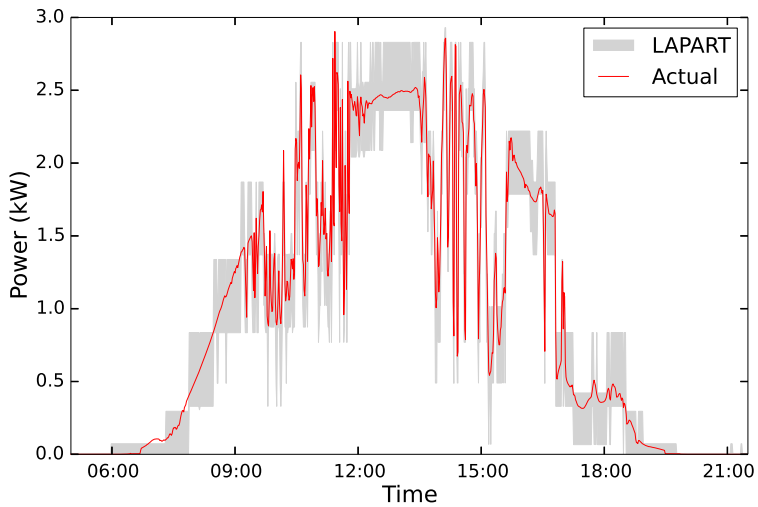
FDD: Classification using LAPART

How does it learn data?





PV Power:

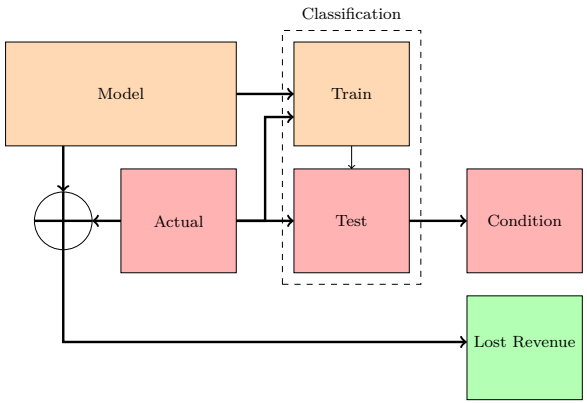


FDD: Classification (Max Power Point Data)

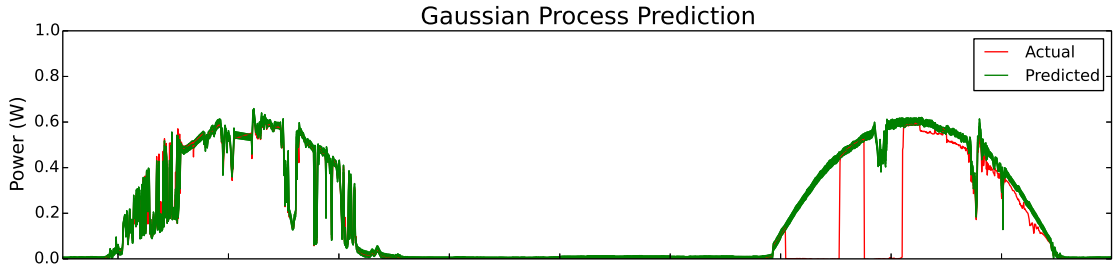
Problem Statement

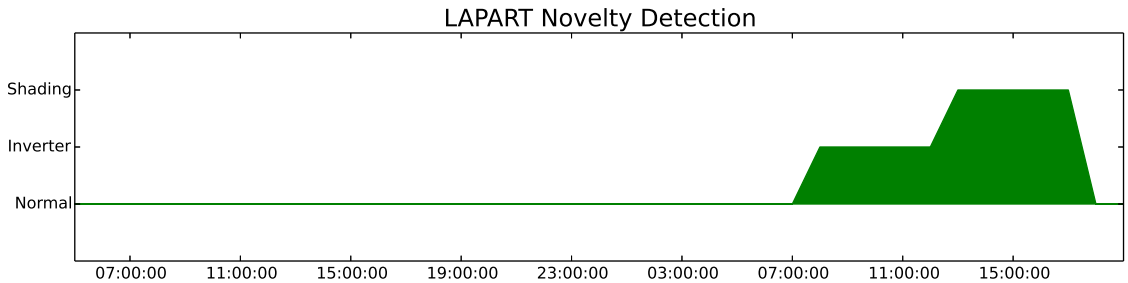
Machine learning can be used to classify fault conditions and estimate lost revenue.

Process



Results





Summary

Machine Learning

1. What?
   - 1.1 machines can learn
   - 1.2 two main types of learning
2. Why?
   - 2.1 Regression
   - 2.2 Classification
3. How?
   - 3.1 Monitor
   - 3.2 Validation
   - 3.3 Faults

Examples

1. Data Quality
   - 1.1 Kernel Density Function (python: scipy, scikit)
2. Performance Validation
   - 2.1 Gaussian, SVM, etc (python: scikit)
3. Novelty Detection/Classification
   - 3.1 Gaussian, SVM, etc (python: scikit)
   - 3.2 LAPART (python: coming soon)

Questions

Sandia National Laboratories

C. Birk Jones
email: cbjones@sandia.gov
phone: 505-844-9261