14 Smart Meters

金力 Li Jin
li.jin@sjtu.edu.cn
上海交通大学密西根学院
Shanghai Jiao Tong University UM Joint Institute

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

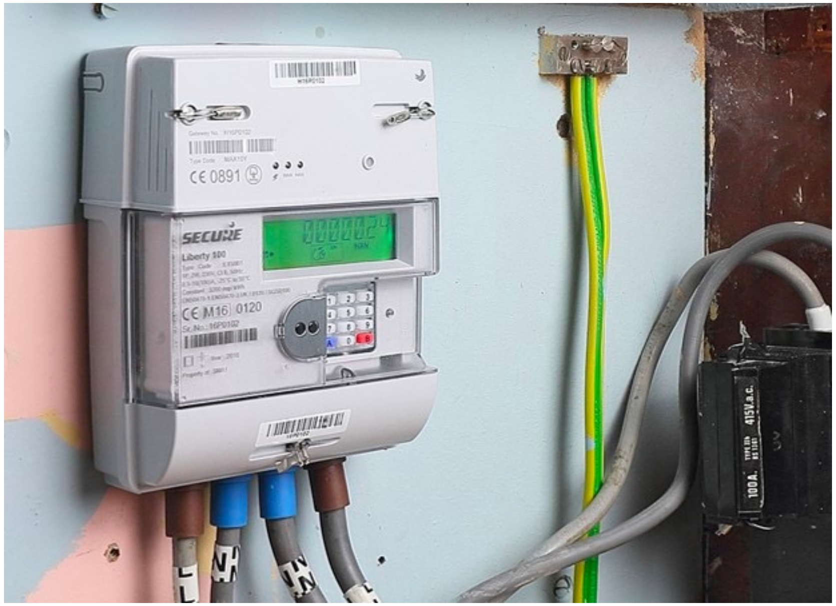
Outline

- Smart meters
- Linear regression & classification
- Analysis results

Smart meters

Electricity load monitoring

- Objective: track use of home appliances
- Motivation: save energy and conduct demand response programs
- Input: load power curve
- Tool: linear classification
- Output: active appliances



Reference

IEEE TRANSACTIONS ON SMART GRID, VOL. 3, NO. 2, JUNE 2012 787

An Event Window Based Load Monitoring Technique for Smart Meters

Ming Dong, *Student Member, IEEE*, Paulo C. M. Meira, *Student Member, IEEE*, Wilsun Xu, *Fellow, IEEE*, and Walmir Freitas, *Member, IEEE*

***Abstract*—The data collected by smart meters contain a lot of useful information. One potential use of the data is to track the energy consumptions and operating statuses of major home appliances. The results will enable homeowners to make sound decisions on how to save energy and how to participate in demand response programs. This paper presents a new method to breakdown the total power demand measured by a smart meter to those used by individual appliances. A unique feature of the proposed method is that it utilizes diverse signatures associated with the entire operating window of an appliance for identification. As a result, appliances with complicated middle process can be tracked. A novel appliance registration device and scheme is also proposed to automate the creation of appliance signature database and to eliminate the need of massive training before identification. The software and system have been developed and deployed to real houses in order to verify the proposed method.**

***Index Terms*—Demand response, load management, load signatures, nonintrusive load monitoring, time-of-use price.**

a data concentrator [4]. While such a sensor network based system can provide accurate measurement of appliance energy consumption, it can be costly and complex to implement. The second direction is to identify and track major home appliances based on the total signal collected by utility meters, which is called nonintrusive load monitoring (NILM) method [5]. Compared to the former, the NILM direction is more attractive to customers and utilities due to its high cost efficiency and less effort on installation.

The problem to be solved by NILM approach can be stated as follows: All the load signals aggregate at the entry point of a house as $P(t)$ and NILM algorithms do the reverse—decode the overall signal into various components $P_i(t)$ that are attributed to specific loads (appliances) i

$$P(t) = P_1(t) + P_2(t) + \cdots + P_n(t). \qquad (1)$$

Background

- The increased public awareness of energy conservation in recent years has created a huge interest in home energy consumption monitoring.
- Consumers show substantial interest in tools that can help them manage their household energy use and expenses.
- A critical link to address this need is the smart meters.
- However, the smart meters currently available in the market can only provide the energy consumption data of a whole house.

Challenge

- Smart meters cannot tell which appliances in the household consume the most energy or are least efficient.
- Also, to take full advantage of time-of-use rates, householders need to be informed of their usage patterns.
- Such information is essential for a household to make sound energy saving decisions and participate in utility demand response programs.

Tow solutions

- In response to this need, two research directions have emerged.
- One is to connect energy monitors to individual appliance of interest and to communicate the recorded data to a data concentrator.
- While such a sensor network-based system can provide accurate measurement of appliance energy consumption, it can be costly and complex to implement.
- The second direction is to identify and track major home appliances based on the total signal collected by utility meters, which is called nonintrusive load monitoring (NILM) method.
- Compared to the former, the NILM direction is more attractive to customers and utilities due to its high cost efficiency and less effort on installation.

Nonintrusive load monitoring

- The problem to be solved by NILM approach can be stated as follows:
- All the load signals aggregate at the entry point of a house as $P(t)$ and NILM algorithms do the reverse—decode the overall signal into various components $P_i(t)$ that are attributed to specific loads (appliances)

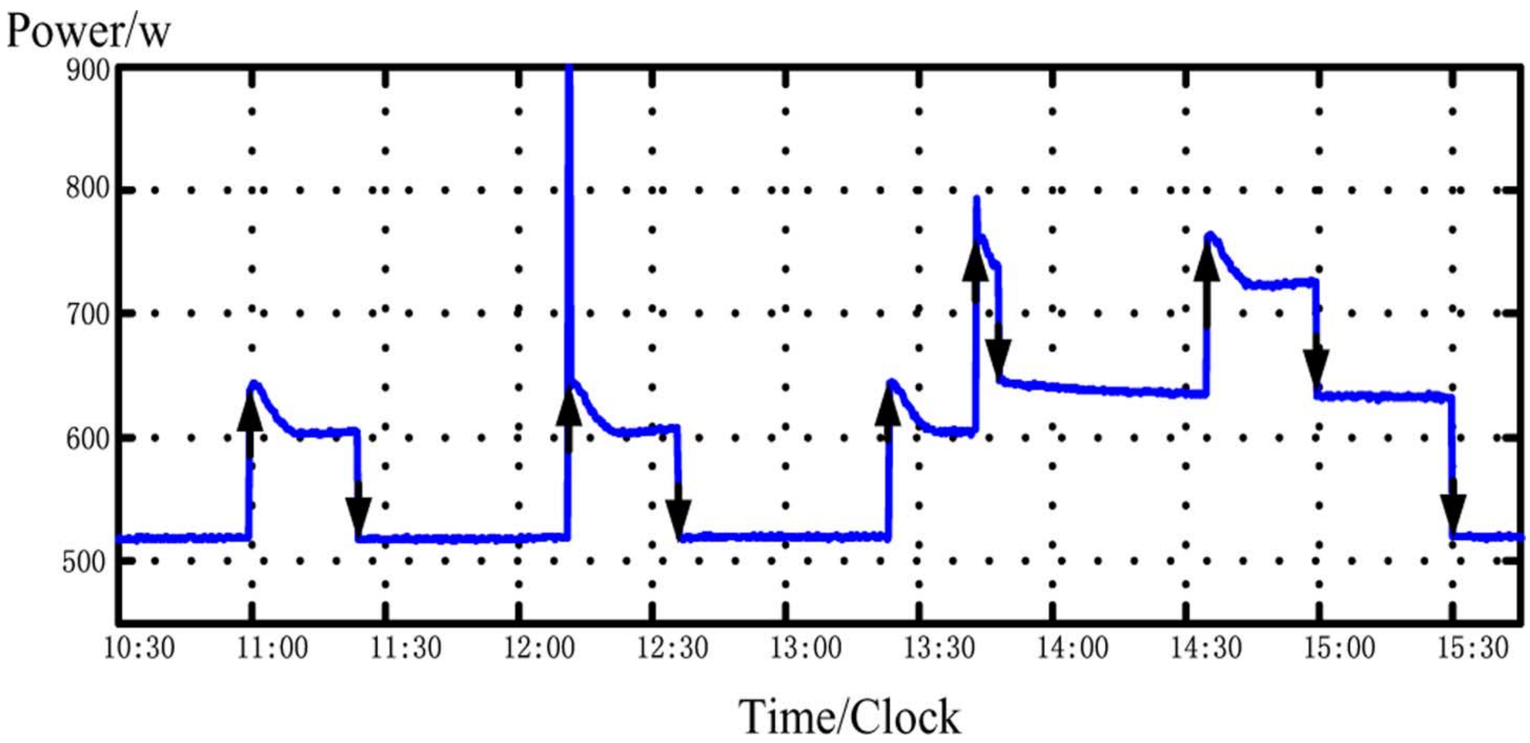
$$P(t) = P_1(t) + P_2(t) + \cdots + P_n(t)$$

- Only large appliances can (and should) be detected; phone chargers etc. are not interesting.
- What are the major power consumers?

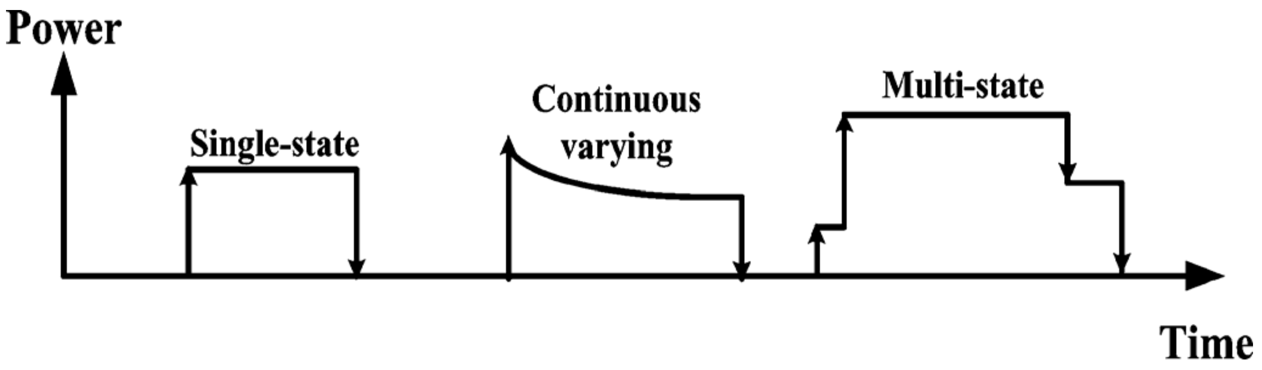
Large appliances

Data

- Appliance events: edges
- Gradual varying power demand between edges



Load types



Load type	Examples	Event	Power demand
Single-state	Light bulb; Toaster	ON=OFF	Flat
Continuous varying	Fridge; Freezer	ON≠OFF	Varying
Multi-state	Furnace; Washer	Multiple events	Varying or flat

Predictors and response

- Input: shape of power curve
  - edge signatures
  - sequence signatures
  - trend signatures
  - time/duration signatures
  - phase signatures
- Output: which appliance is active
  - We can have multiple appliances ON

Classification problem

- Like most engineering problems, the most difficult part is setting up the model.
- That is also the most effort-consuming part.
- Specifically, generating predictors is the main challenge
  - One the one hand, they should capture what intuitively make sense
  - On the other hand, they are in formats compatible with classification algorithms
- A typical paradigm: human generate the predictors, while computer trains the model.

Linear regression & classification

Hypothetical linear relation

- Linear regression hypothesizes that response variable linearly varies with independent variable, i.e.
$$y = \beta_0 + \beta_1 x + e$$
- β_0 is the intercept, β_1 is the coefficient, e is a zero-mean random variable called the noise
- Noise = randomness due to our lack of information
- We assume that e follows normal distribution
- This is a hypothetical relation, which may or may not be true

Fitting a straight line

- Suppose that n repeated experiments have been done
- The experiments use $x_1, x_2, \ldots, x_n$ as the values of the dependent variables.
- The experiments lead to the following results: $y_1, y_2, \ldots, y_n$.
- Hypothesize that $y = \beta_0 + \beta_1 x$
- What should β_0 and β_1 be? -> least squares

Method of least squares*

- Given β_0 and β_1, the prediction error of the linear model $y = \beta_0 + \beta_1 x$ is
$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$
- Principle of method of least squares: find parameters β_0 and β_1 that minimize the sum of prediction errors
$$(\hat{\beta}_0, \hat{\beta}_1) = \arg\min S(\beta_0, \beta_1)$$
- Least square!

Formulae for univariate LR*

- See notes for derivation.

$$\hat{\beta}_0 = \frac{\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$\hat{\beta}_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

Multivariate linear regression

- Consider $p-1$ independent variables $x_1, x_2, \dots, x_{p-1}$
- One dependent variable y
- Hypothetical relation

$$y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_{p-1} x_{p-1} + e$$

- To determine the intercept and coefficients, minimize the square error

$$S = \sum_{i=1}^{n} \left(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{i,1} + \cdots + \hat{\beta}_{p-1} x_{i,p-1}) \right)^2$$

Multivariate linear regression

- Consider $p-1$ independent variables $x_1, x_2, \dots, x_{p-1}$
- One dependent variable y
- In matrix form

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \ \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1,p-1} \\ 1 & x_{21} & x_{22} & \cdots & x_{2,p-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{n,p-1} \end{bmatrix}$$

- Be careful: X begins with a column of 1's!
- Hypothetical relation

$$Y = X\beta$$

- So we want to minimize square erro (SE) $S(\beta) = \|Y - X\beta\|^2$
- $\|v\|^2 = v_1^2 + v_2^2 + \cdots + v_m^2$

Formulae for model parameters

- The formulae of least square estimation:

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

- See notes for derivation
- To implement this, you may want to refresh your knowledge about
  - Matrix multiplication
  - Transpose
  - Inverse
  - How to implement them in Excel/Python/MATLAB/c++…
  - For example, in Excel we have mmult and minverse

Expected value of $\boldsymbol{\beta}$*

THEOREM A

Under the assumptions of the standard statistical model, the least squares estimates are unbiased: $E(\hat{\beta}_j) = \beta_j$, for $j = 0, 1$.

Proof

From the assumptions, $E(y_i) = \beta_0 + \beta_1 x_i$. Thus, from the equation for $\hat{\beta}_0$ in Section 14.1,

$$
\begin{aligned}
E(\hat{\beta}_0) &= \frac{\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} E(y_i)\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i E(y_i)\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \\
&= \frac{\left(\sum_{i=1}^{n} x_i^2\right)\left(n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \\
&= \beta_0
\end{aligned}
$$

The proof for β_1 is similar. ■

Variance of $\boldsymbol{\beta}$*

THEOREM **B**

Under the assumptions of the standard statistical model,

$$\mathrm{Var}(\hat{\beta}_0) = \frac{\sigma^2 \sum_{i=1}^{n} x_i^2}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$\mathrm{Var}(\hat{\beta}_1) = \frac{n\sigma^2}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$\mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\sigma^2 \sum_{i=1}^{n} x_i}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

Variance of $\boldsymbol{\beta}*$

Proof

From a form for $\hat{\beta}_1$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

The identity for the numerator follows from expanding the product and using $\sum(x_i - \bar{x}) = 0$. We then have

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

which reduces to the desired expression. The other expressions may be derived similarly. ■

Residual sum of squares (RSS)*

$$\text{RSS} = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

- Quantifies the deviation between true response and predicted response.
- Recall hypothetical linear relation
$$Y = X\beta + E$$
- $E = [e_1, e_2, \ldots, e_n]^T$
- $e_1, e_2, \ldots, e_n$ are independent and identically distributed (IID) random variables following normal distribution with mean 0 and variance σ^2
- An unbiased estimation for σ^2: $s^2 = \dfrac{\text{RSS}}{n-2}$

Confidence interval*

- We can prove that

$$\frac{\hat{\beta}_i - \beta_i}{s_{\hat{\beta}_i}} \sim t_{n-2}$$

- $s_{\widehat{\beta}_i}$ is the estimate of β_i with RSS instead of σ^2
- t_{n-2}: student's t distribution with degree of freedom n-2
- Can be used to generate 95% confidence interval

Check how well linear regression fits data

- Use correlation of determination (R-squared)
- Total sum of squares; $\bar{y}$ = average value of $y_1, y_2, \dots, y_n$

$$S_Y = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

- Residual sum of squares

$$RSS = \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 x_{i,1} - \cdots - \beta_{p-1} x_{i,p-1}\right)^2$$

- Coefficient of determination

$$R^2 = 1 - \frac{RSS}{S_Y}$$

- $R^2 = 1$: perfectly linear. $R^2 = 0$: not linear

Variations of linear regression

- Suppose that $y = \beta_0 + \beta_1 f(x)$
- $f(x)$ can be a nonlinear function of x
- Then, we can fit a model by minimizing

$$RSS = \sum \left(y_i - \beta_0 - \beta_1 f(x_i)\right)^2$$

- For example, the nonlinear function can be
  - $f(x) = x^2$
  - $f(x) = \log x$
  - $f(x) = e^x$
  - $f(x) = \sin x$

Linear classification

- Predictors: $x = (x_1, x_2, \ldots, x_p)$
- k features of a sample
- Response: $G \in \{1,2, \ldots, K\}$
- Every sample belongs to one out of K classes
- Classification problem: given predictors $X = x$, what is the probability that this sample belongs to a particular class?
$$\Pr\{G = k|X = x\} =?$$
- If we can compute this probability, then our prediction of class is
$$\hat{G} = \arg\max_k \Pr\{G = k|X = x\}$$

Linear regression of an indicator matrix

- Predictors: $x = (x_1, x_2, \dots, x_p)$
- Response: $y = (y_1, y_2, \dots, y_K)$
  - If an observation is in class k, then $y_k = 1$ and $y_j = 0$ for $j \neq k$
- Assume that the indicator variables $y_1, y_2, \dots, y_K$ linearly depend on the predictors
  - $y_1 = \beta_{1,0} + \beta_{1,1}x_1 + \beta_{1,2}x_2 + \cdots + \beta_{1,p}x_p$
  - $y_2 = \beta_{2,0} + \beta_{2,1}x_1 + \beta_{2,2}x_2 + \cdots + \beta_{2,p}x_p$
  - …
  - $y_K = \beta_{K,0} + \beta_{K,1}x_1 + \beta_{K,2}x_2 + \cdots + \beta_{K,p}x_p$
- In matrix form
  - $y = X^T B$
  - y is $K \times 1$, B is $(p+1) \times K$, X is $(p+1) \times 1$

Linear regression of an indicator matrix

- How to determine coefficient matrix B?
- Suppose that we have n observations
- RSS = $\sum_{i=1}^{n}\sum_{k=1}^{K}(y_{i,k} - \beta_{k,0} + \beta_{k,1}x_1 + \beta_{k,2}x_2 + \cdots + \beta_{k,p}x_p)$ = $(Y - X^T B)^T(Y - X^T B)$
- Minimize RSS

$$\hat{B} = (X^T X)^{-1} X^T Y$$

- Fitted classes

$$\hat{Y} = X^T \hat{B} = X(X^T X)^{-1} X^T Y$$

Prediction

A new observation with input x is classified as follows

1. Compute the fitted output $\hat{f}(x) = (1, x^T)\hat{B}$, a K vector
2. Identify the largest component and classify accordingly

$$\hat{G}(x) = \arg\max_k \hat{f}_k(x)$$

Why?

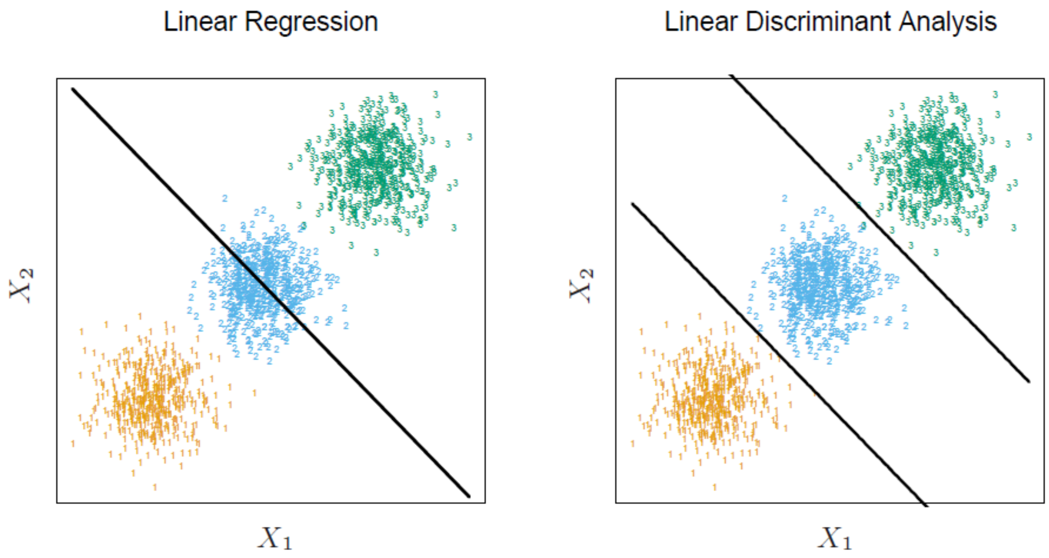
- A rather formal justification: y_k is a random variable such that

$$E[\hat{y}_k | X = x] = \Pr\{G = k | X = x\}$$

- That is, $\hat{y}_k \approx \Pr\{G = k | X = x\}$
- Illustration for two-class problem

LR not always work

- It seems that $\hat{y}_k \approx \Pr\{G = k|X = x\}$
- However, $\hat{y}_k = \beta_{k,0} + \beta_{k,1}x_1 + \beta_{k,2}x_2 + \cdots + \beta_{k,p}x_p$ can be < 0 or > 1, which is not allowed for probabilities
- Another serious problem: masking
- Illustration for three-class problem



Linear discriminant analysis*

- Consider N observations
- Predictor vector x
- Class g
- Define
  - $\hat{\pi}_k = N_k/N$, where N_k is the number of class-k observations
  - $\hat{\mu}_k = \sum_{g_i=k} x_i/N_k$
  - $\hat{\Sigma} = \sum_{k=1}^{K} \sum_{i:g_i=k}(x_i - \hat{\mu}_k)(x_i - \hat{\mu}_i)^T/(N-K)$
- Linear discriminant function

$$\delta_k(x) = x^T\Sigma^{-1}\mu_k - \frac{1}{2}\mu_k^T\Sigma^{-1}\mu_k + \log \pi_k$$

- Classification

$$G(x) = \arg\max_k \delta_k(x)$$

Logistic regression*

- Logit function

$$\Pr\{G = k | X = x\} = \frac{\exp\left(\beta_{k0} + \beta_k^T x\right)}{\sum_{l=1}^{K} \exp(\beta_{l0} + \beta_l^T x)}$$

- Classification

$$G(x) = \arg\max_k \frac{\exp\left(\beta_{k0} + \beta_k^T x\right)}{\sum_{l=1}^{K} \exp(\beta_{l0} + \beta_l^T x)}$$

- Process of fitting coefficients β_{ki}: called logistic regression
- Use maximum likelihood

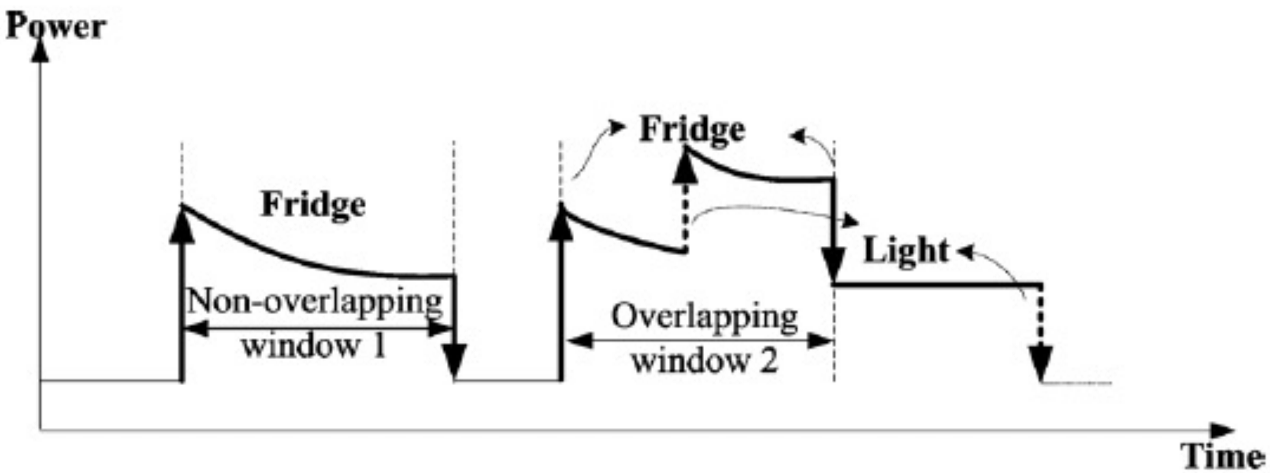
Summary of linear classification*

- Predictor x
- Find some function $f_k(x)$ such that
$$f_k(x) \approx \Pr\{G = k|X = x\}$$
- For a new observation with input x, classify as follows
$$\hat{G}(x) = \arg\max_k f_k(x)$$
- $f_k(x)$ is linear in x -> linear regression
- $f_k(x) = x^T\Sigma^{-1}\mu_k - \frac{1}{2}\mu_k^T\Sigma^{-1}\mu_k + \log\pi_k$ -> LDA
- $f_k(x) = \frac{\exp(\beta_{k0}+\beta_k^T x)}{\sum_{l=1}^K \exp(\beta_{l0}+\beta_l^T x)}$ -> logistic regression

Back to smart meters

Event windows

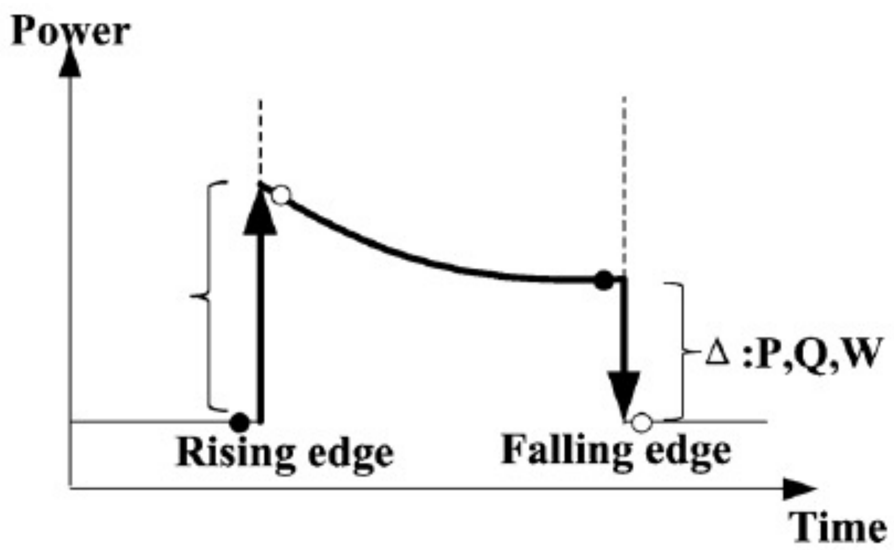
- An event window is defined as the collection of all signatures between any pair of rising/falling step-changes (events) of the power demand as measured by the smart meter.



- Overlapping: most appliances
- Non-overlapping: short-duration or always-on

Edge signature

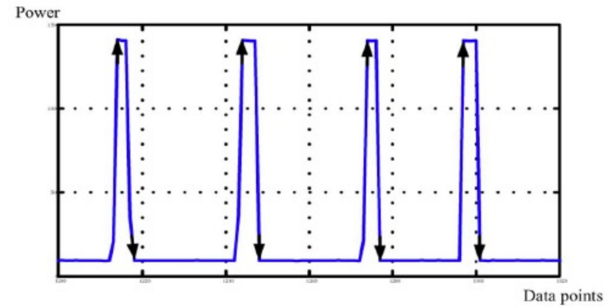
- An edge refers to the event of the operating state of an appliance, which be a step change in its power demand.
- The edge can be either rising or falling.
- Each edge can be characterized by the changes in power (P), reactive power (Q), and current waveform (W)



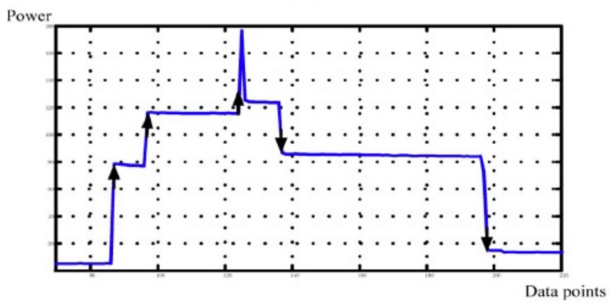
Sequence signature

- Sequence signature describes the logical sequence of operation events of a load.
- In another word, it represents the sequence of appearances of edges.
- For example, a washer usually follows the following operating modes: water-fill, immerse, rinse, drainage, and spin-dry.

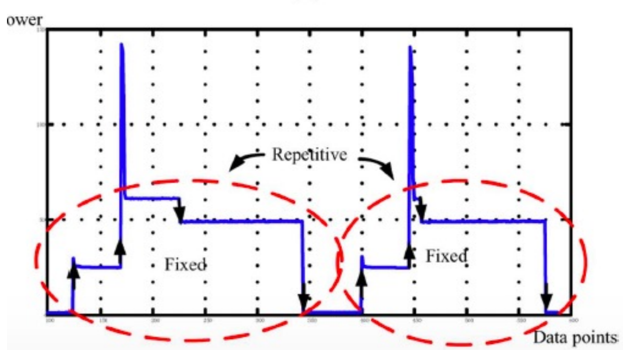
Sequence signature



(a)



(b)

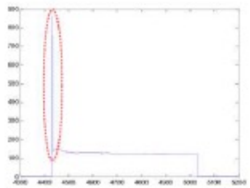
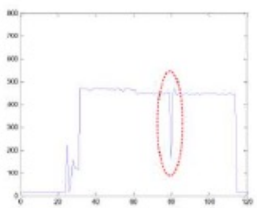
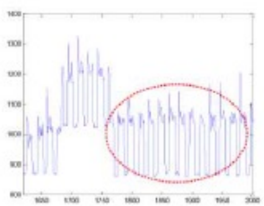
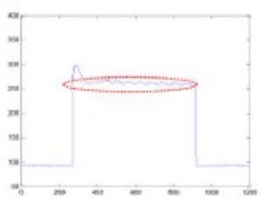
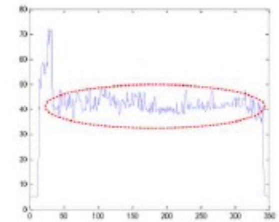
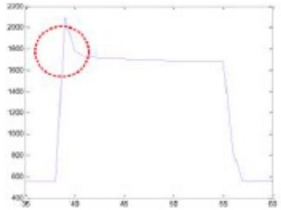
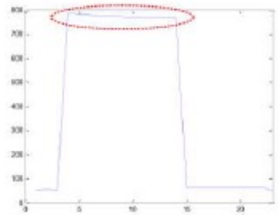


(c)

Trend signature

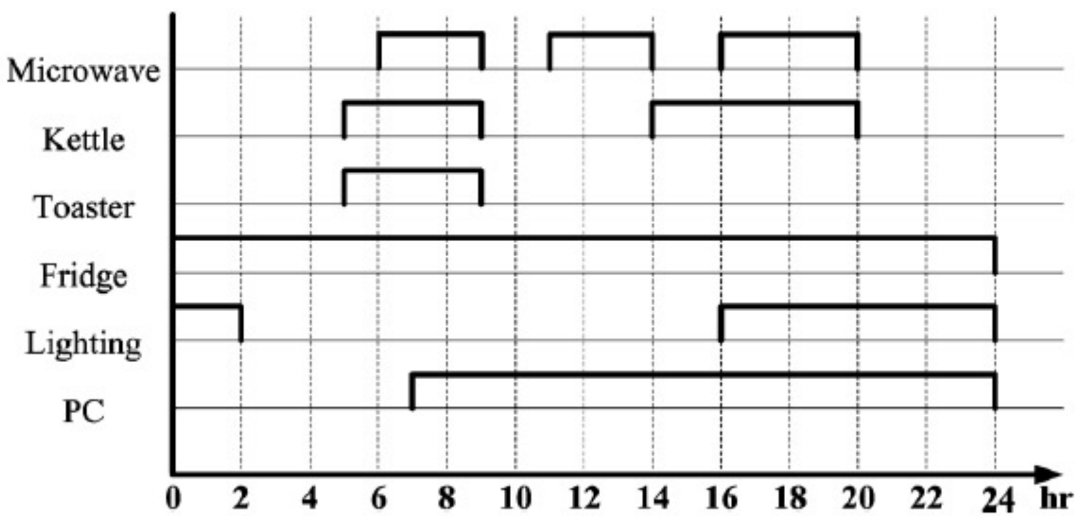
- A trend signature refers to variation of power demand between two edges.
- A TV set may experience a falling spike at moments of switching channels; pulses are usually caused by electronic switches.
- A lot of stoves have pulses because they have an integer-cycle controller in it; it prevents itself from overheating.

Trend signature

Type	Curve example	Power slope feature
Rising spike		A large negative slope following a larger positive slope
Falling spike		A large positive slope following a large negative slope
Pulses		Continuous pairs of large slopes
Fluctuation		Continuous small slopes; signs of slopes slowly change
Quick vibrate		Continuous small slopes; signs of slopes quickly change
Gradual falling		Continuous small negative slopes
Flat		Continuous small slopes

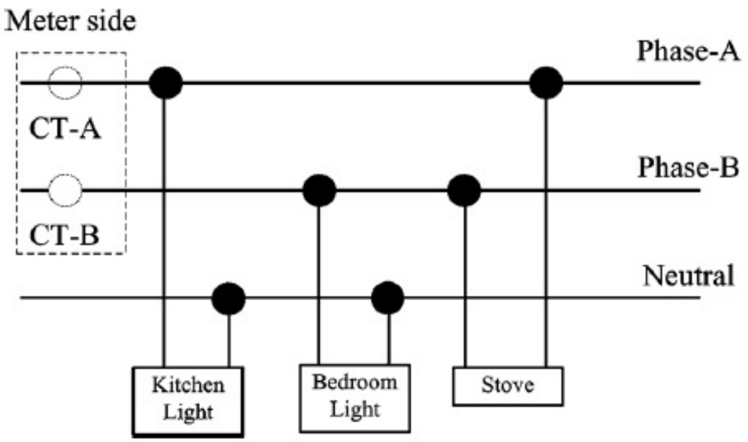
Time/Duration signature

- The time of load window appearance relates close to its function.
- Microwaves are more expected to be seen before breakfast, lunch, and supper
- Lights are usually turned on in the early morning or after dark
- Fridge and furnace are likely to run throughout 24 h.



Phase signature

- There are two 120 V hot wires installed in a typical North American residential house.
- Hereby, the two wires can be named as A and B. Most appliances are connected between A or B and neutral.
- However, some heavy appliances such as stove and dryer are connected between A and B to gain a 240 V voltage.
- We can differentiate from the phase pattern.



Load identification

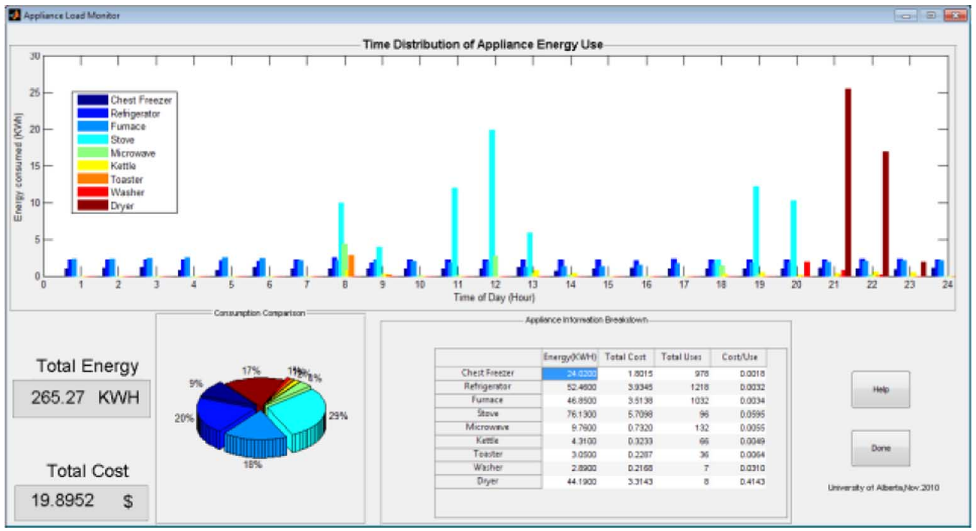
- Inputs x vector
  - S_{edge}
  - S_{seq}
  - S_{trd}
  - S_{time}
  - S_{phase}
- Classifier $G_k(x) = \beta^T x - \gamma_k$
- γ_k is a threshold for type k appliance
- Type k is ON if $G_k(x) > 0$
- β^T and γ_k determined by linear regression and optimized by validation set

Results

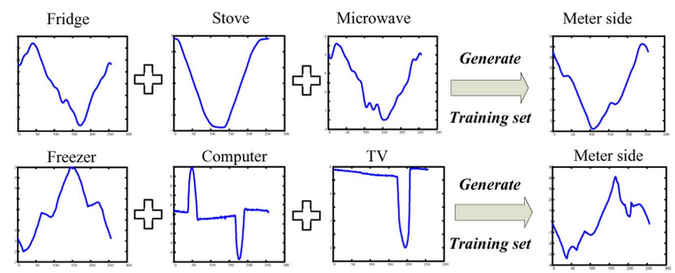
- Weights associated with first four signatures

Load name	Distinctive signatures	ω^T	γ
Fridge	Edge, trend	[0.53 0.17 0.22 0.08]	0.85
Microwave	Edge, time	[0.65 0.19 0 0.16]	0.85
Furnace	Edge, sequence	[0.53 0.47 0 0]	0.85
Stove	Edge, sequence, time	[0.51 0.3 0 0.19]	0.85
Washer	Edge, sequence	[0.55 0.45 0 0]	0.8
Kettle	Edge	[0.86 0.14 0 0]	0.85
Laptop	Edge, trend	[0.5 0.25 0.25 0]	0.8
Average	---	[0.59 0.28 0.07 0.06]	0.85

Verification



Appliance Name	Actual operation times	Correctly identified operation times	False identified operation times	Identification accuracy(%)
Chest Freezer	178	163	2	90.5
Fridge	213	203	4	93.4
Furnace	185	172	1	92.4
Stove/Oven	16	16	0	100
Microwave	23	22	0	95.7
Kettle	12	11	0	91.6
Toaster	6	6	0	100
Washer	3	3	0	100
Dryer	4	4	0	100



Summary

- What are linear regression & linear classification?
- What are the inputs & outputs to the smart meter problem?
- Why linear regression cannot be used to multiple (more than two) classes?
- How to convert power consumption data to input of a linear classification model?