05 | Introduction to Machine Learning

Cynthia Rudin | MIT Sloan School of Management

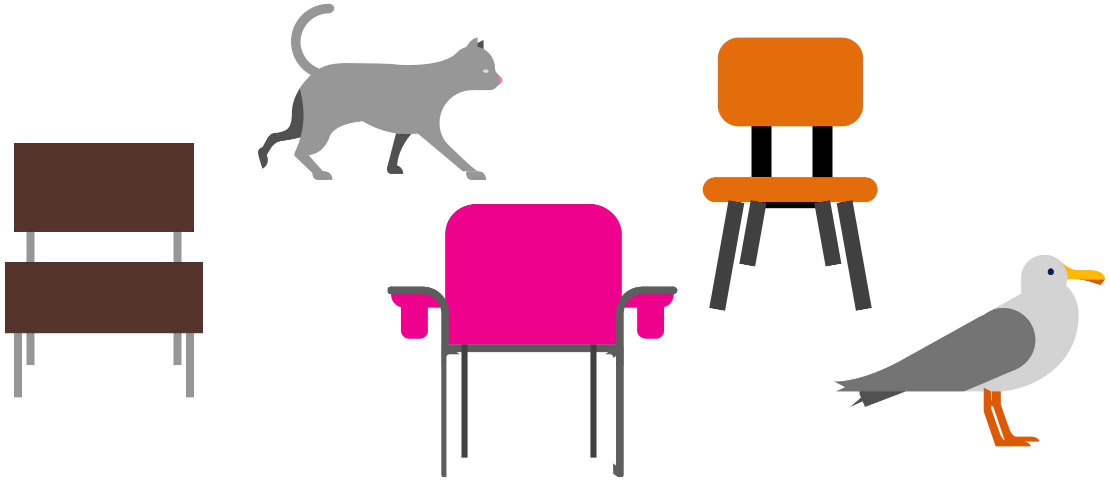
Introduction to Machine Learning

- Classification – Predict answers to Yes/No questions
- Regression – Predict real values
- Clustering – Find patterns of similar objects
- How to Evaluate Machine Learning Models

Classification

Machine Learning

- Grew out of artificial intelligence within computer science. Teaches computers by example.

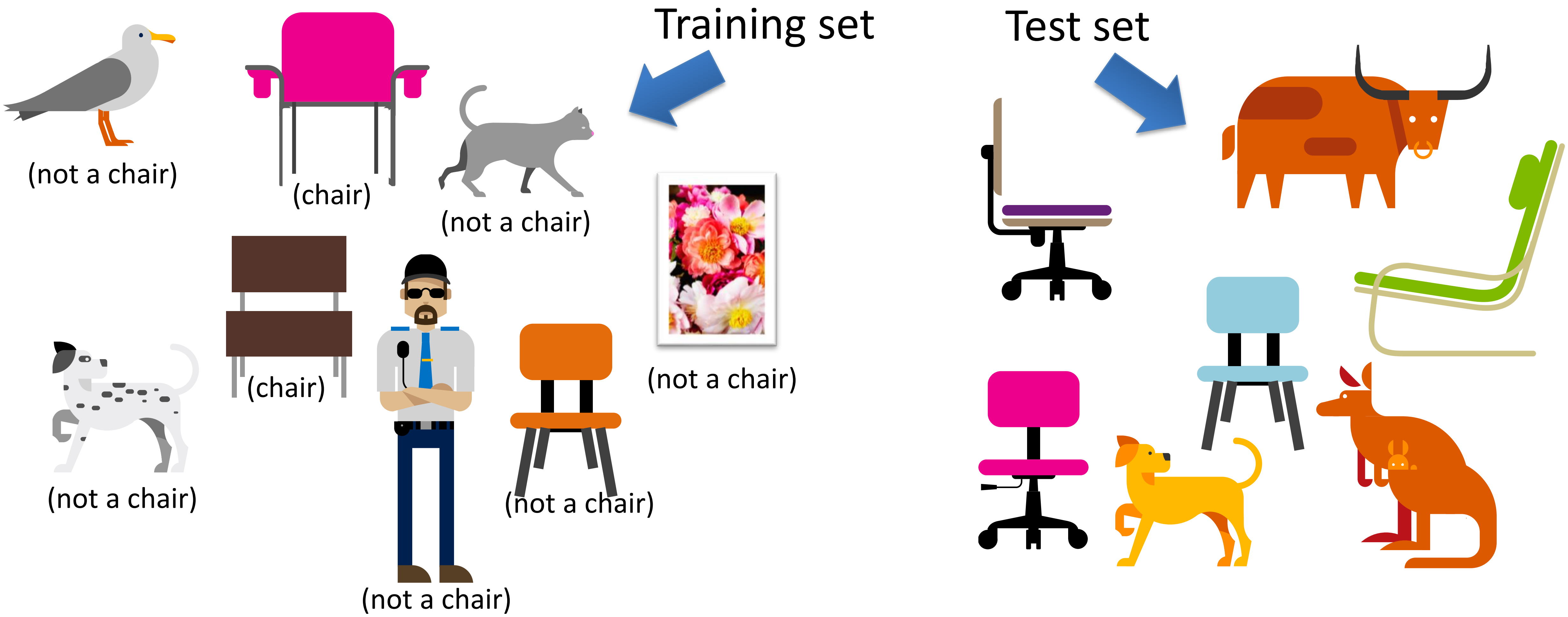


Is this a chair?

- ML is arguably part of statistics.

Classification

We have a *training set* of observations (e.g., labeled images) and a *test set* that we use only for evaluation.



Classification

- Each observation is represented by a set of numbers (features).

Each pixel gets rgb values like [1.0,0.9,0.8]

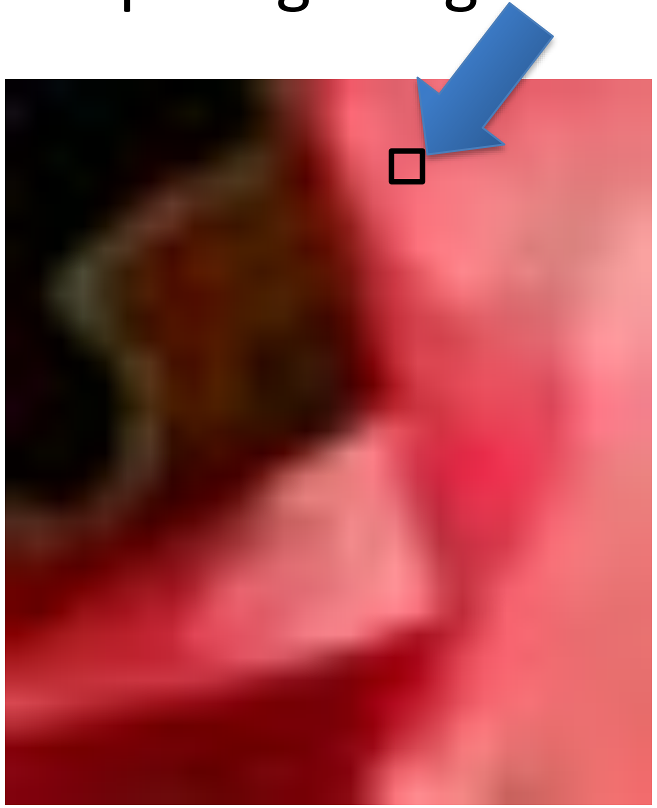


Image becomes:
[1.0,0.9,0.8,0.1,0.5,...]

(Label is -1, it's not a chair)

Classification

- Each observation is represented by a set of numbers (features).

Manhole is represented as: [5 3 120 12 1 0 …..]

Number of events last year
Number of serious events last year
Number of electrical cables
Number of pre-1930 electrical cables
Vented cover?
Inspected?

Classification

- Each observation is represented by a set of numbers (features).

Manhole is represented as: [5 3 120 12 1 0 …..]

Number of events last year

Number of serious events last year

Number of electrical cables

Number of pre-1930 electrical cables

Vented cover?

Inspected?

Training feature data is from 2014 and before

Classification

- Each observation is represented by a set of numbers (features).

Manhole is represented as: [5 3 120 12 1 0 …..]

Number of events last year
Number of serious events last year
Number of electrical cables
Number of pre-1930 electrical cables
Vented cover?
Inspected?

Training feature data is from 2014 and before
Label is 1 if it had an event in 2015

Classification

- Each observation is represented by a set of numbers (features).

Manhole is represented as: [5 3 120 12 1 0 …..]

Number of events last year

Number of serious events last year

Number of electrical cables

Number of pre-1930 electrical cables

Vented cover?

Inspected?

Testing feature data is from 2015 and before

Predict what happen in 2016

Classification

- Each observation is represented by a set of numbers (features).

Manhole is represented as:	[5	3	120	12	1	0	.....	]	-1
	[0	0	89	5	1	1	.....	]	1
	[1	0	20	0	0	1	.....	]	-1
:			:						

Features, called X

Labels, called Y

(Predictors, Covariates,
Explanatory Variables,
Independent Variables)

Classification

- Formally, given training set $(x_{i,} y_i)$ for i=1…n, we want to create a classification model f that can predict label y for a new x.

Classification

- Formally, given training set $(x_{i,}y_i)$ for i=1…n, we want to create a classification model f that can predict label y for a new x.

Manhole is represented as: [1925 15]

Year oldest cable installed

Number of events last year

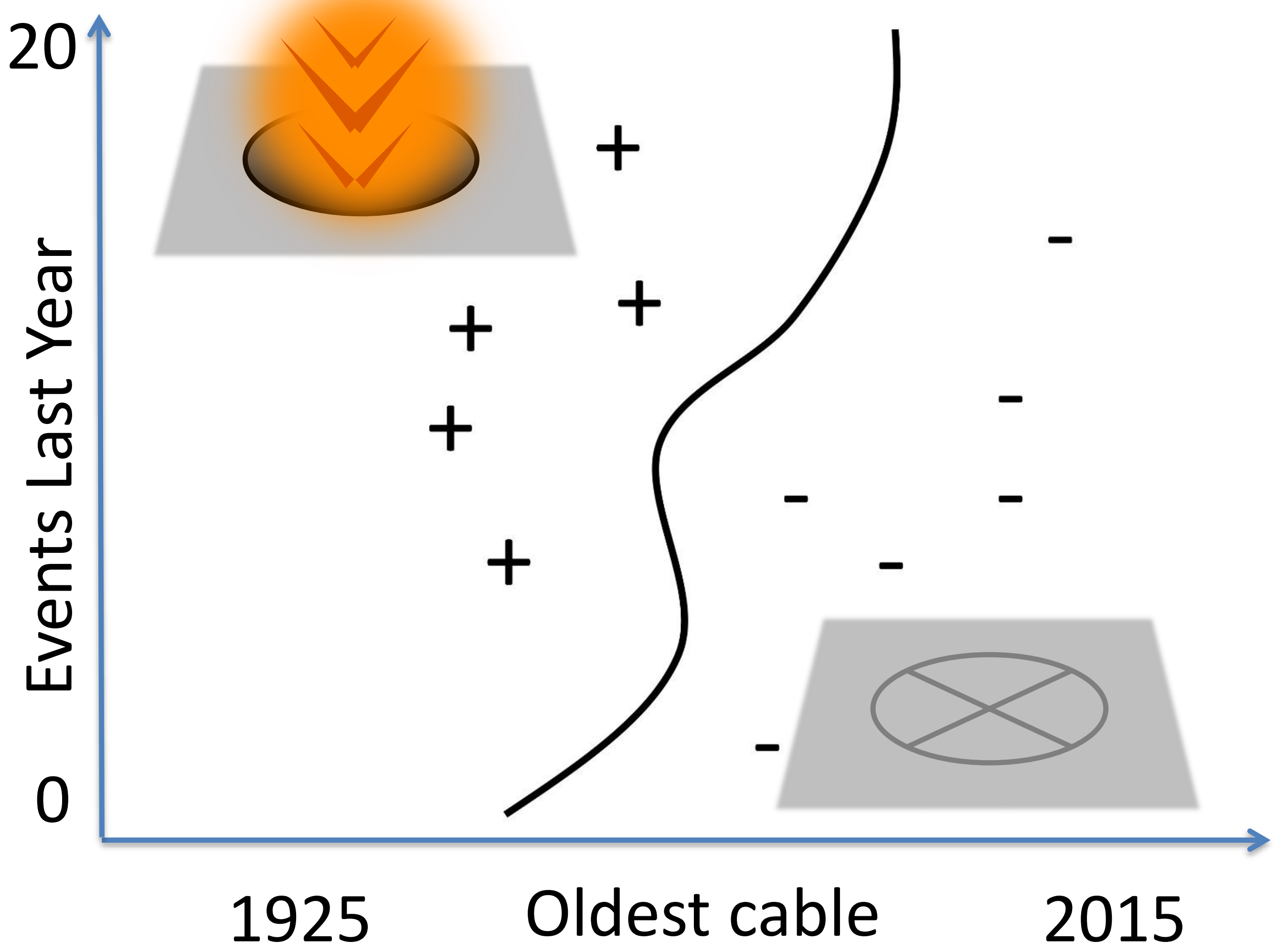
Classification

- Formally, given training set (x_i, y_i) for i=1…n, we want to create a classification model f that can predict label y for a new x.

Manhole is represented as: [1925 15]

Year oldest cable installed

Number of events last year



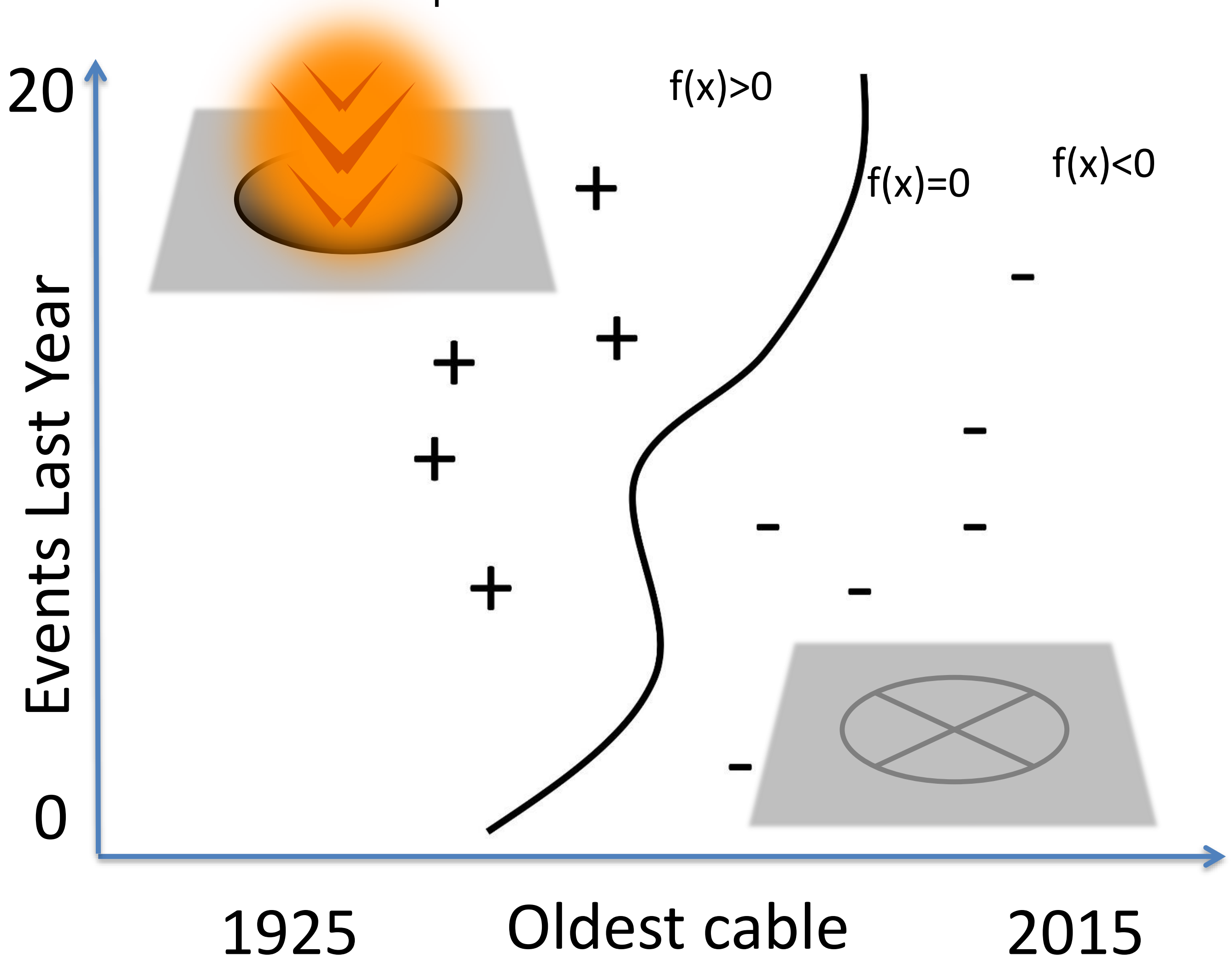
Classification

- Formally, given training set (x_i, y_i) for i=1…n, we want to create a classification model f that can predict label y for a new x.

Manhole is represented as: [1925   15]

Year oldest cable installed

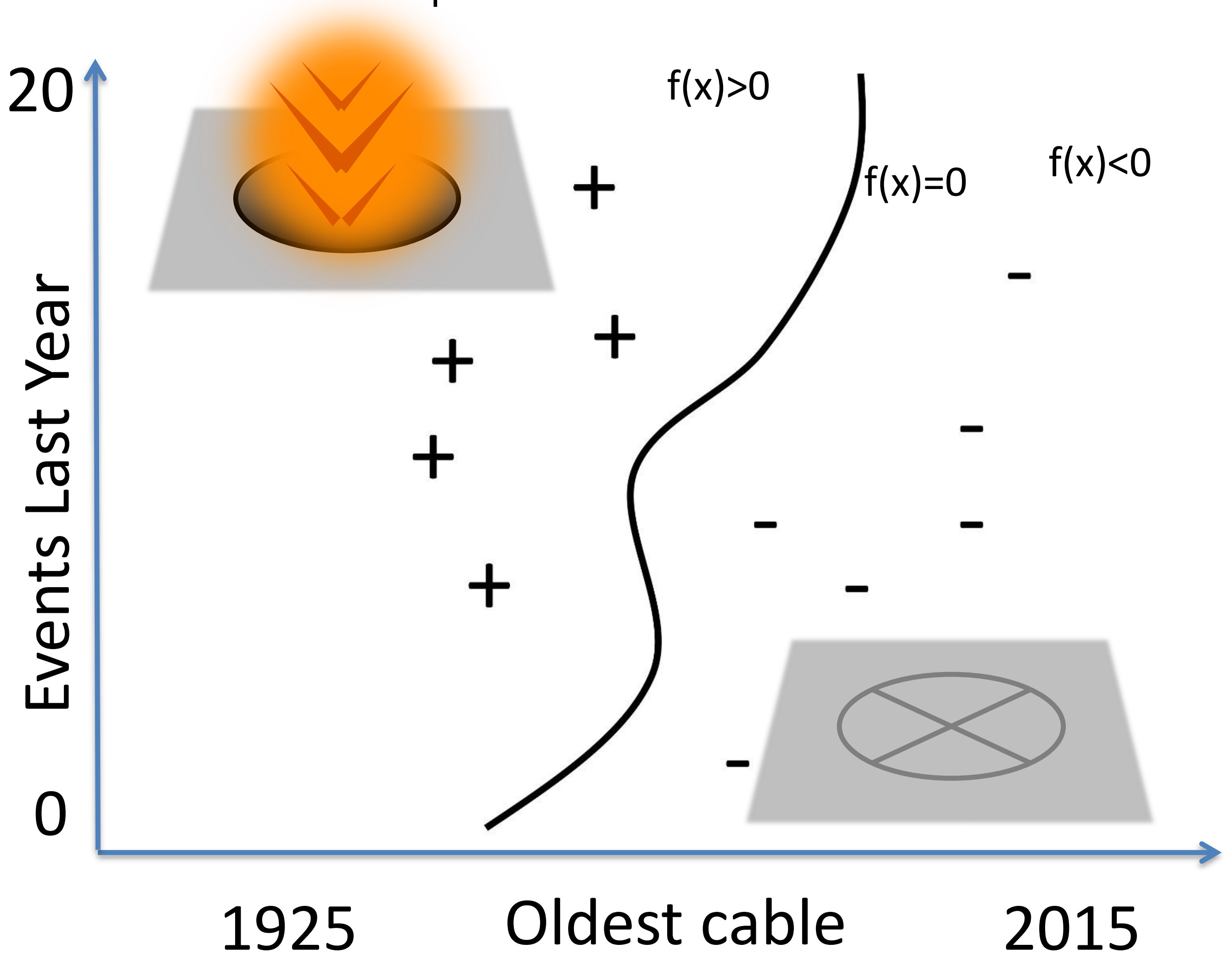
Number of events last year



Classification

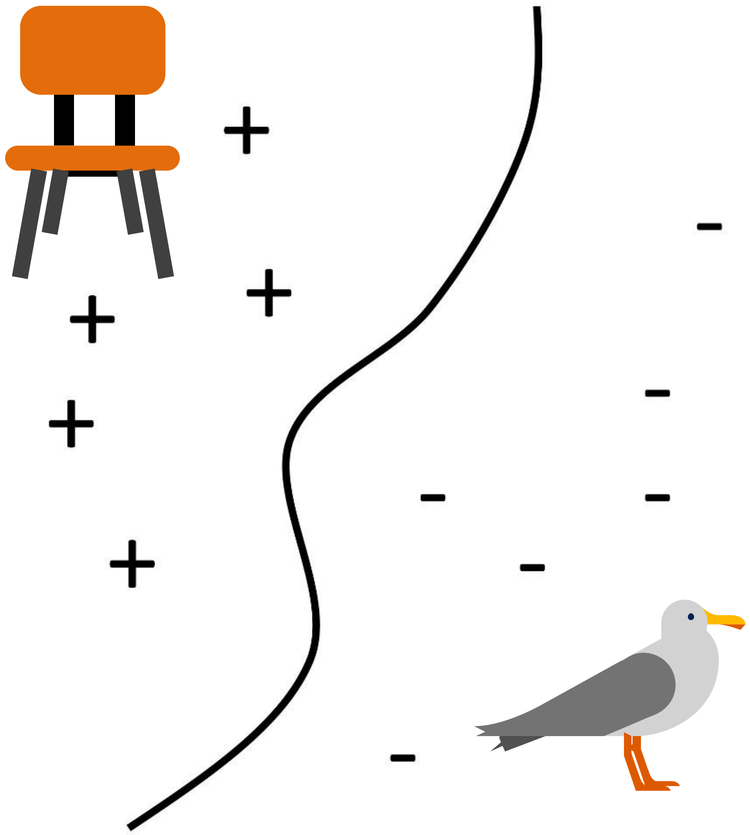
- Formally, given training set (x_i, y_i) for i=1…n, we want to create a classification model f that can predict label y for a new x.

f(x) = function(Events Last Year, Oldest Cable)



Classification

- Formally, given training set (x_i, y_i) for i=1…n, we want to create a classification model f that can predict label y for a new x.



Classification

- Formally, given training set (x_i, y_i) for i=1…n, we want to create a classification model f that can predict label y for a new x.

- The machine learning algorithm will create the function f.

- The predicted value of y for a new x is sign(f(x)).

Classification

- Yes/No questions – binary classification
- automatic handwriting recognition, speech recognition, biometrics, document classification, spam detection, predicting credit default risk, detecting credit card fraud, predicting customer churn, predicting medical outcomes (strokes, side effects, etc.)

Classification

- Common algorithms:
  - Logistic Regression (with L1 or L2 regularization)
  - Decision Trees / Classification Trees / CART / C4.5 / C5.0
  - AdaBoost (Boosted Decision Trees)
  - Support Vector Machines
  - Random Forests
  - Neural Networks
- You never need to program these.

Regression

Regression

- For predicting real-valued outcomes:
  - How many customers will arrive at our website next week?
  - How many tv’s will we sell next year?
  - Can we predict someone’s income from their click through information?

Regression

- Each observation is represented by a set of numbers.

Income

A person is represented as: [5 3 120 12 1 0 .....] 84

[0 0 89 5 1 1 .....] 32

[1 0 20 0 0 1 .....] -10

: :

Features, called X

Labels, called Y

Regression

- Each observation is represented by a set of numbers.

A person is represented as:	[5]	84
	[0]	32
	[1]	-10
	:	

Income

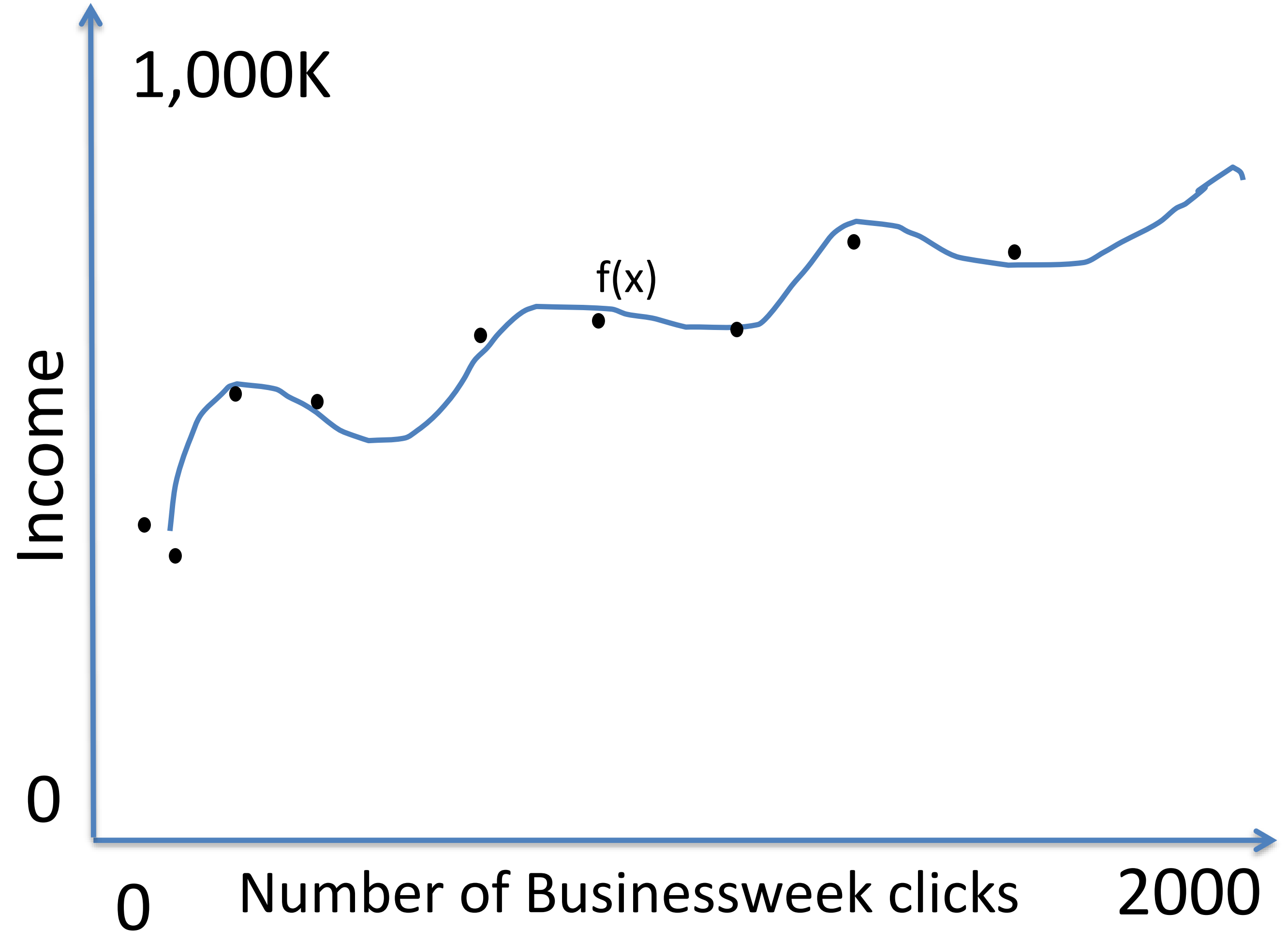
Single feature, called X

Labels, called Y

Regression

- Formally, given training set (x_i, y_i) for i=1…n, we want to create a regression model f that can predict label y for a new x.

f(x) = function(Number of Businessweek clicks)

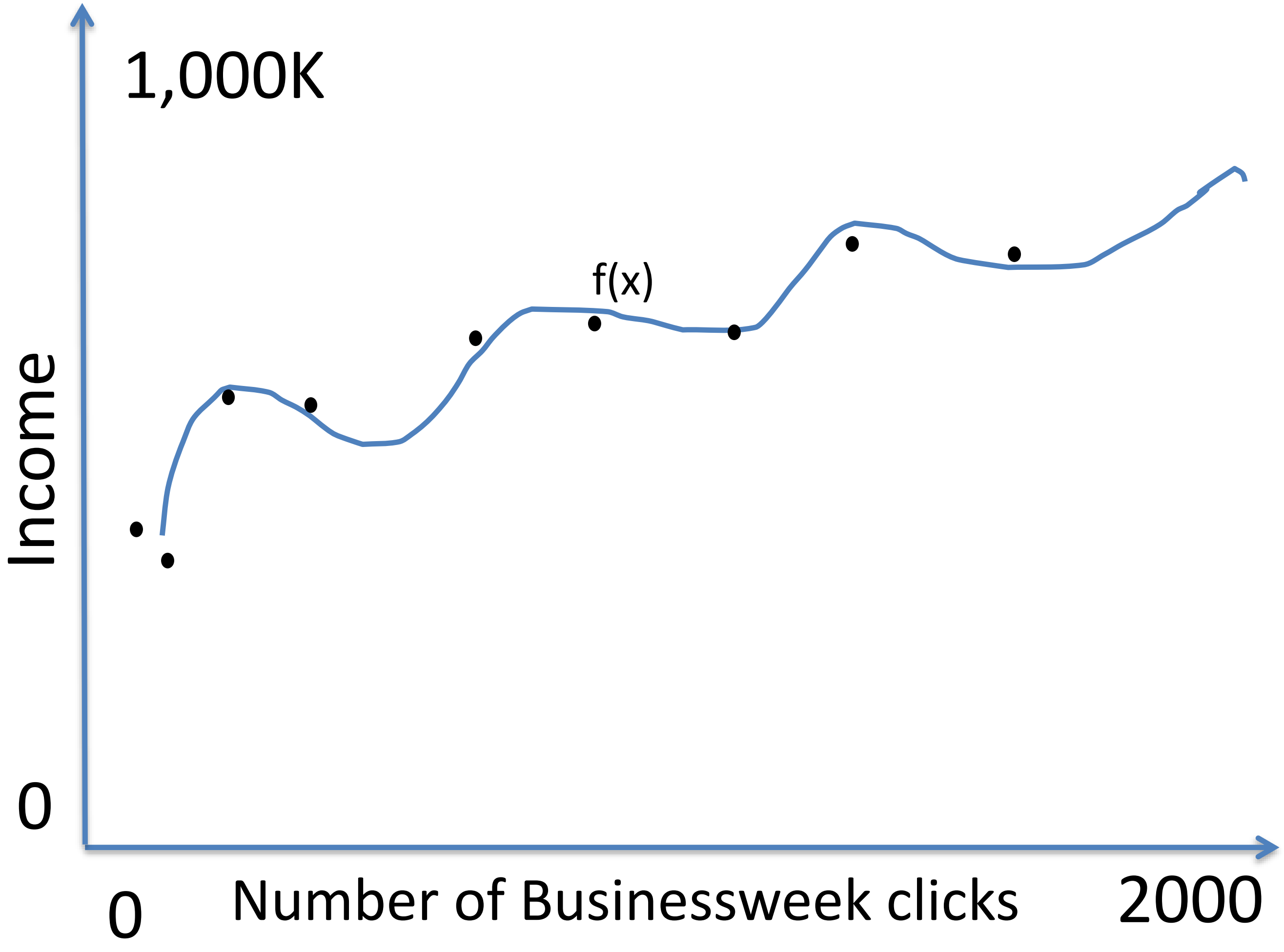


Regression

- Formally, given training set (x_i, y_i) for i=1…n, we want to create a regression model f that can predict label y for a new x.

f(x) = function(Number of Businessweek clicks)

(Overfitting?)

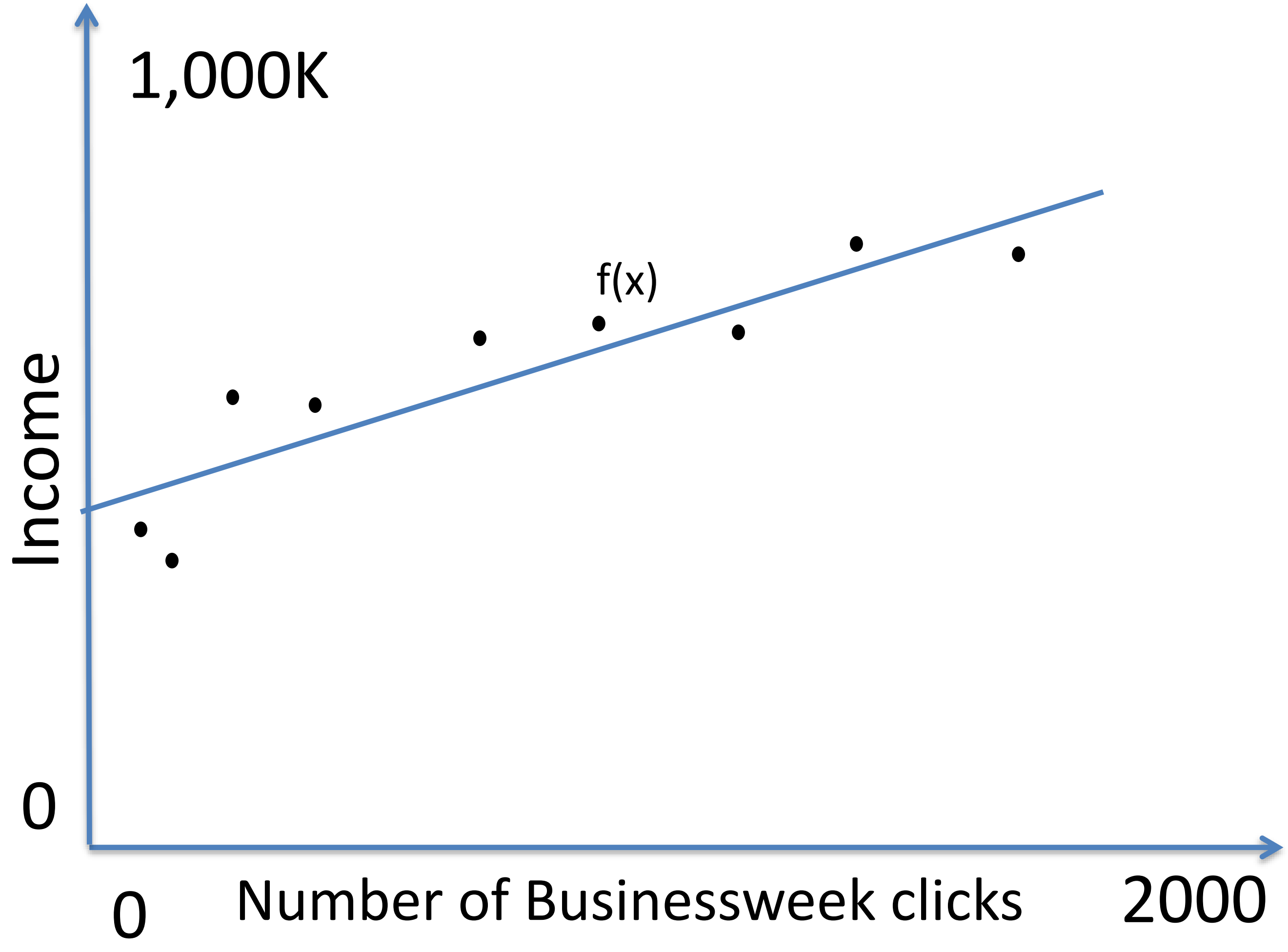


Regression

- Formally, given training set (x_i, y_i) for i=1…n, we want to create a regression model f that can predict label y for a new x.

f(x) = function(Number of Businessweek clicks)
= 5K*Number of Businessweek clicks + 100K

(Underfitting?)



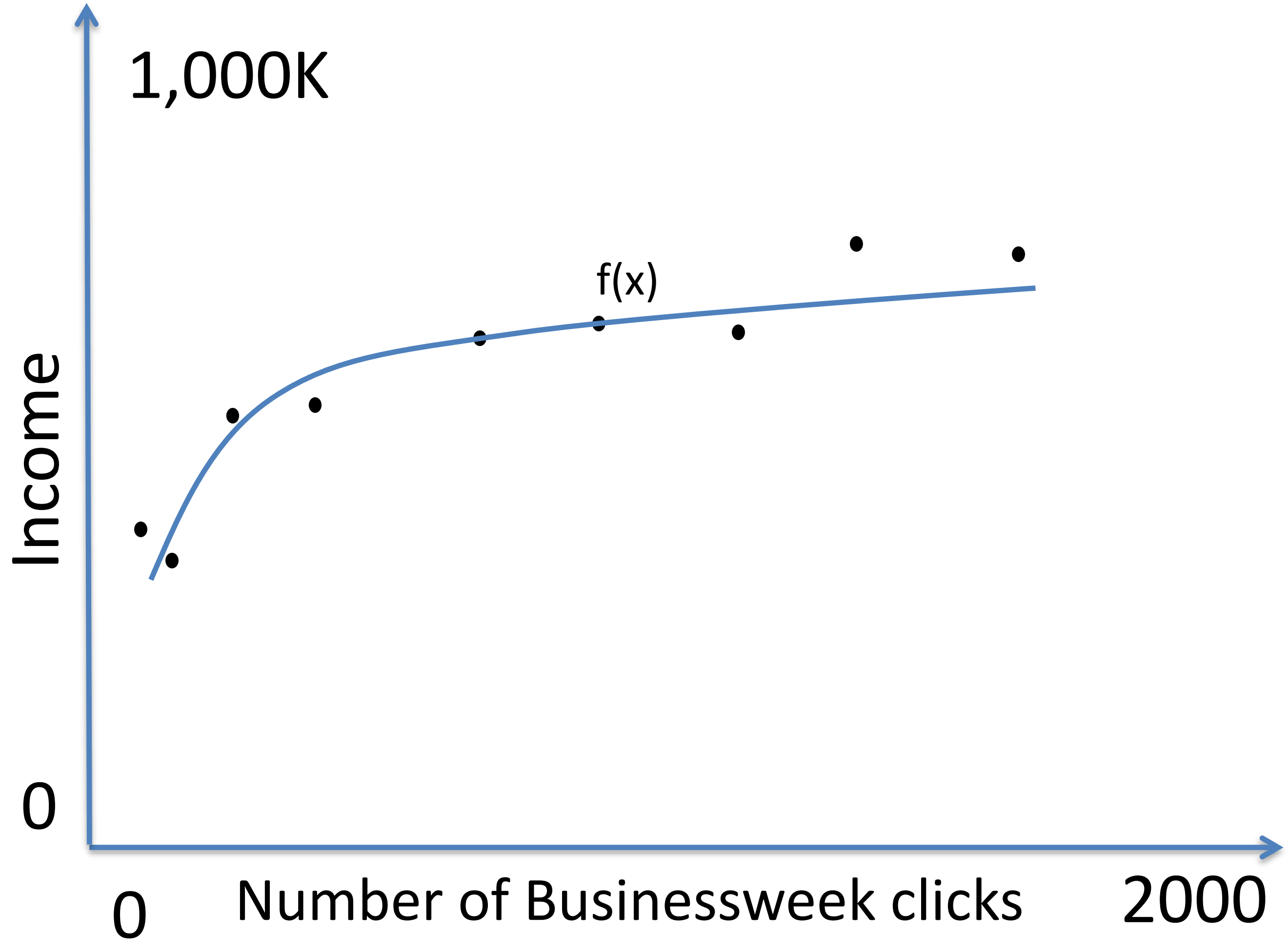
Regression

- Formally, given training set (x_i, y_i) for i=1…n, we want to create a regression model f that can predict label y for a new x.

f(x) = function(Number of Businessweek clicks)

(Just right?)

We'll talk more about this later



Regression

- Formally, given training set (x_i, y_i) for i=1…n, we want to create a regression model f that can predict label y for a new x.

Estimated income:

f(x) = function(Number of visits to upscale furniture websites, Number of Businessweek clicks, Number of distinct people emailed per day, Number of purchases of over 5K within the last month, Number of visits to airlines, etc.)

For instance,
f(x) = 3*Number of visits to upscale furniture websites
+10*Number of Businessweek clicks
+100*Number of distinct people emailed per day
+2*Number of purchases of over 5K within the last month
+10*Number of visits to airlines

But f(x) could be much more complicated

Regression Applications

- Predict monetary amounts
- Predict consumption or demand for products/energy

Supervised Learning

- “Supervised” means that the training data has ground truth labels to learn from. Classification and Regression are supervised learning problems.
- (Supervised) classification often has +1 or -1 labels.
- (Supervised) regression has numerical labels.
- Supervised learning algorithms are much easier to evaluate than unsupervised ones.

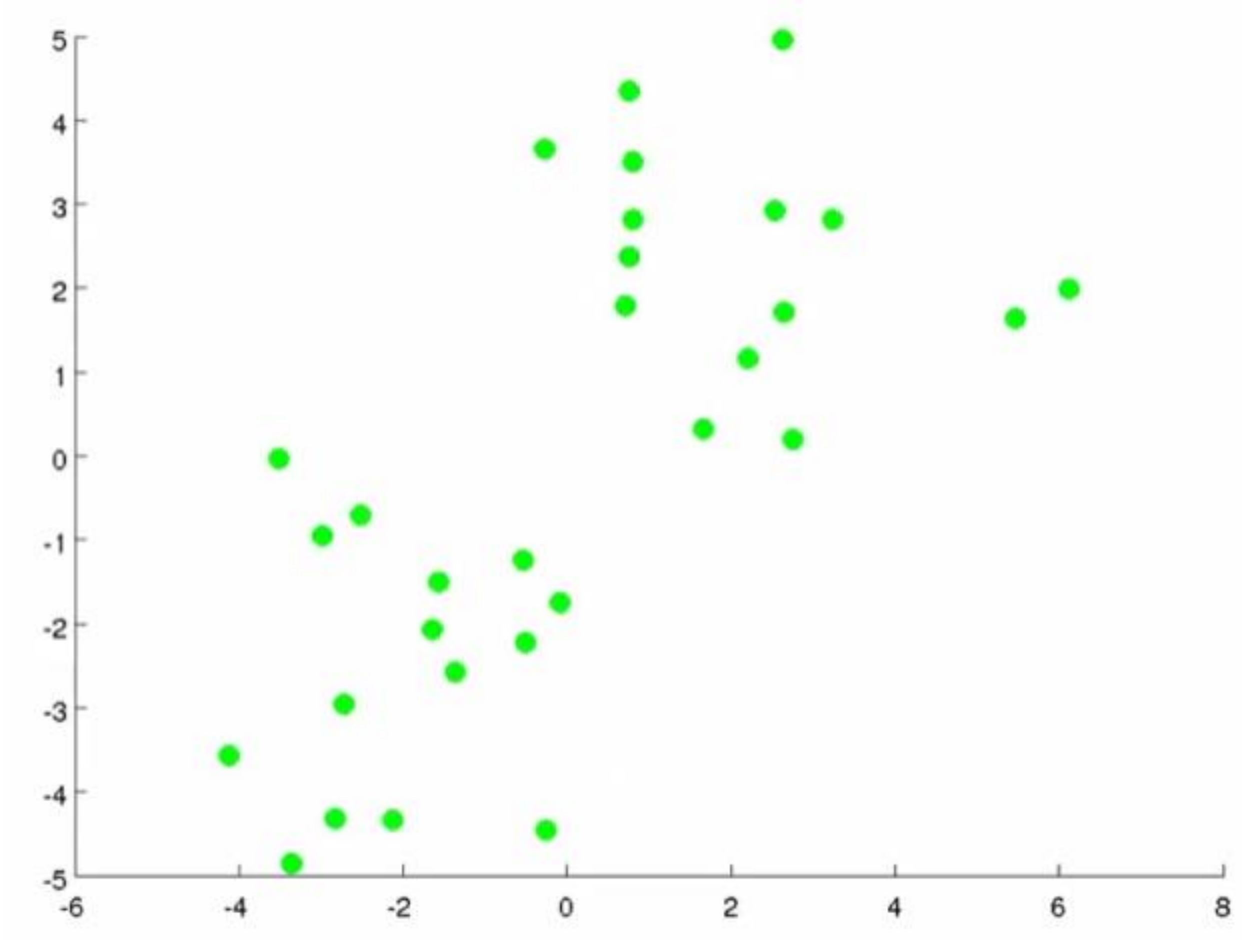
Clustering

Microsoft

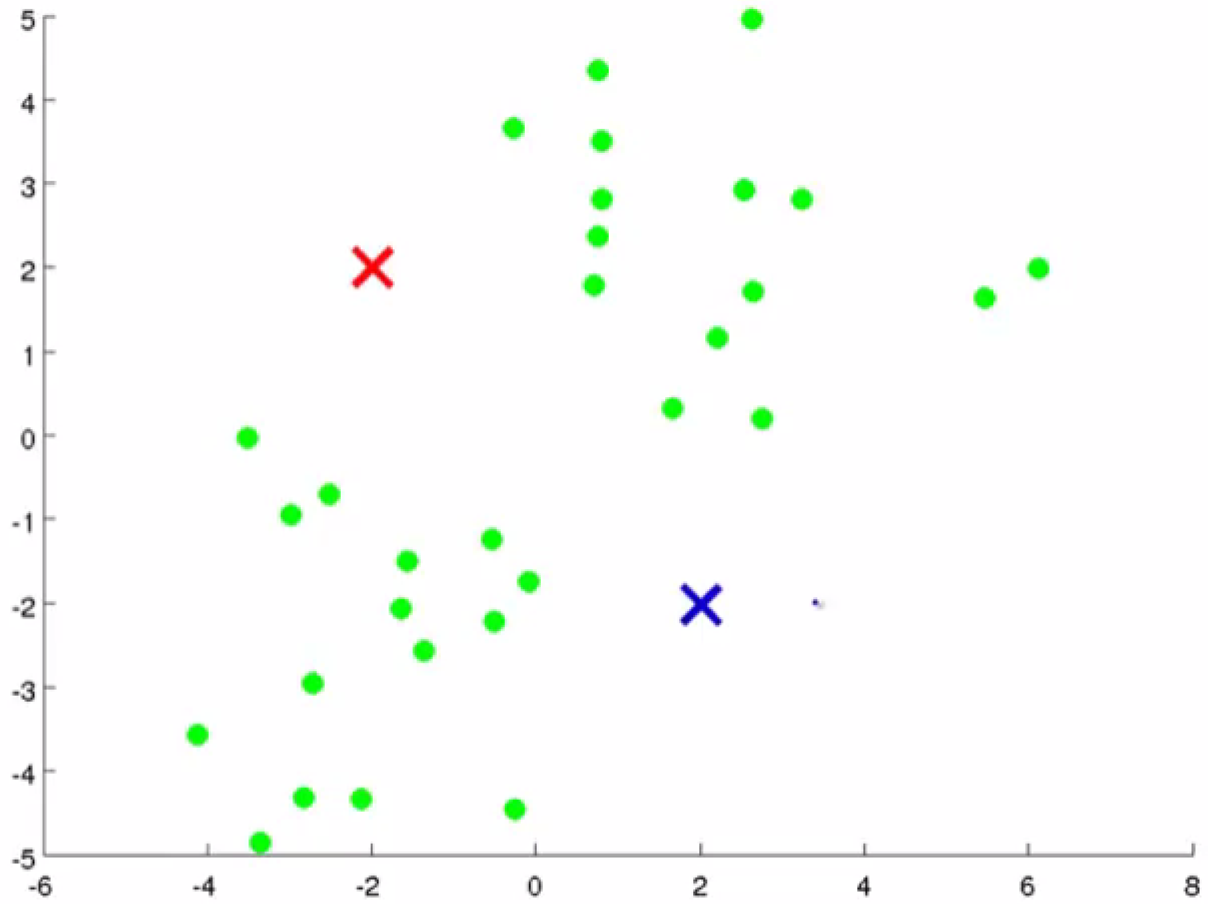
K-means

- It is a method of grouping, which aims to partition a set of n observations into k groups, such that each observation belongs to the group closest to the mean.

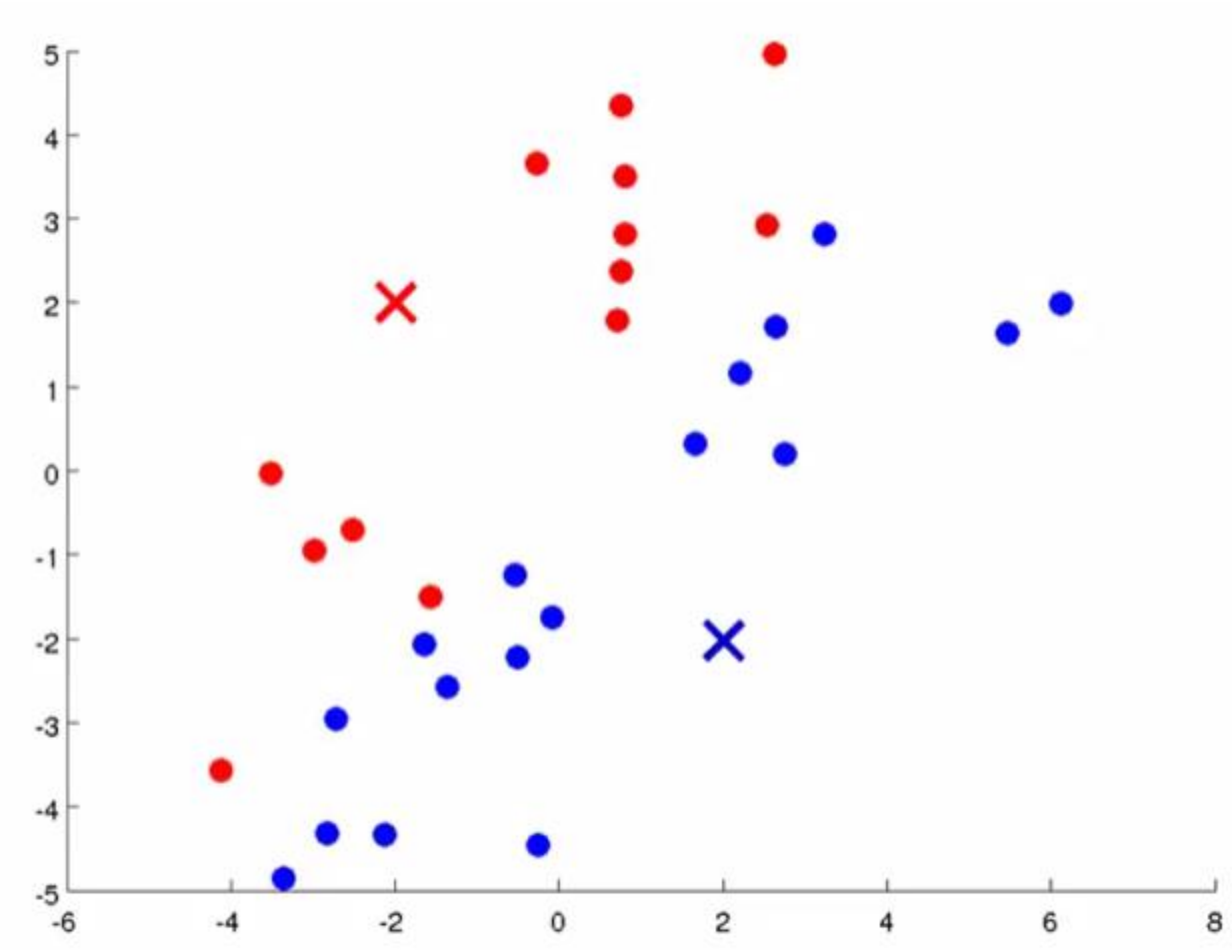
K-medias



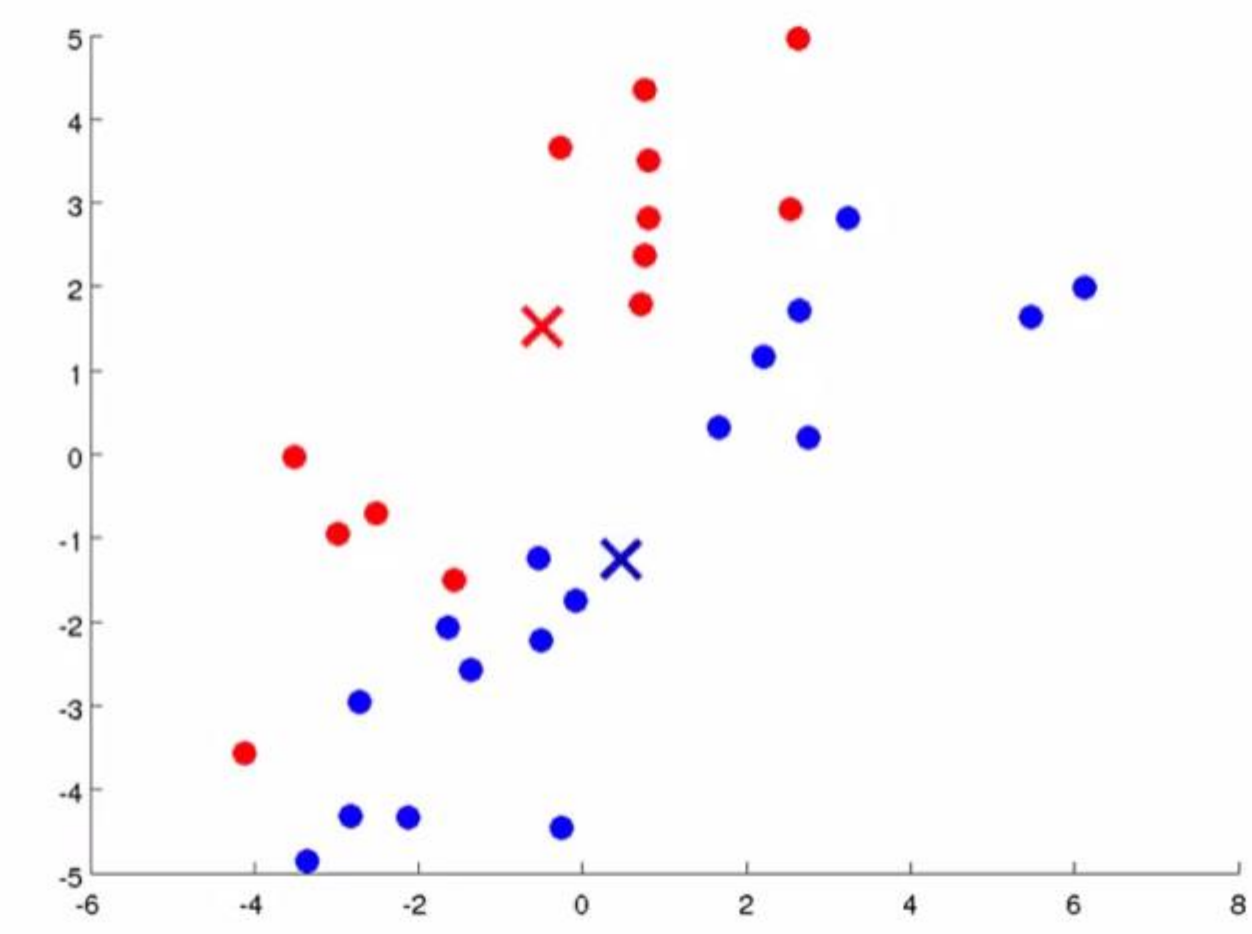
K-medias



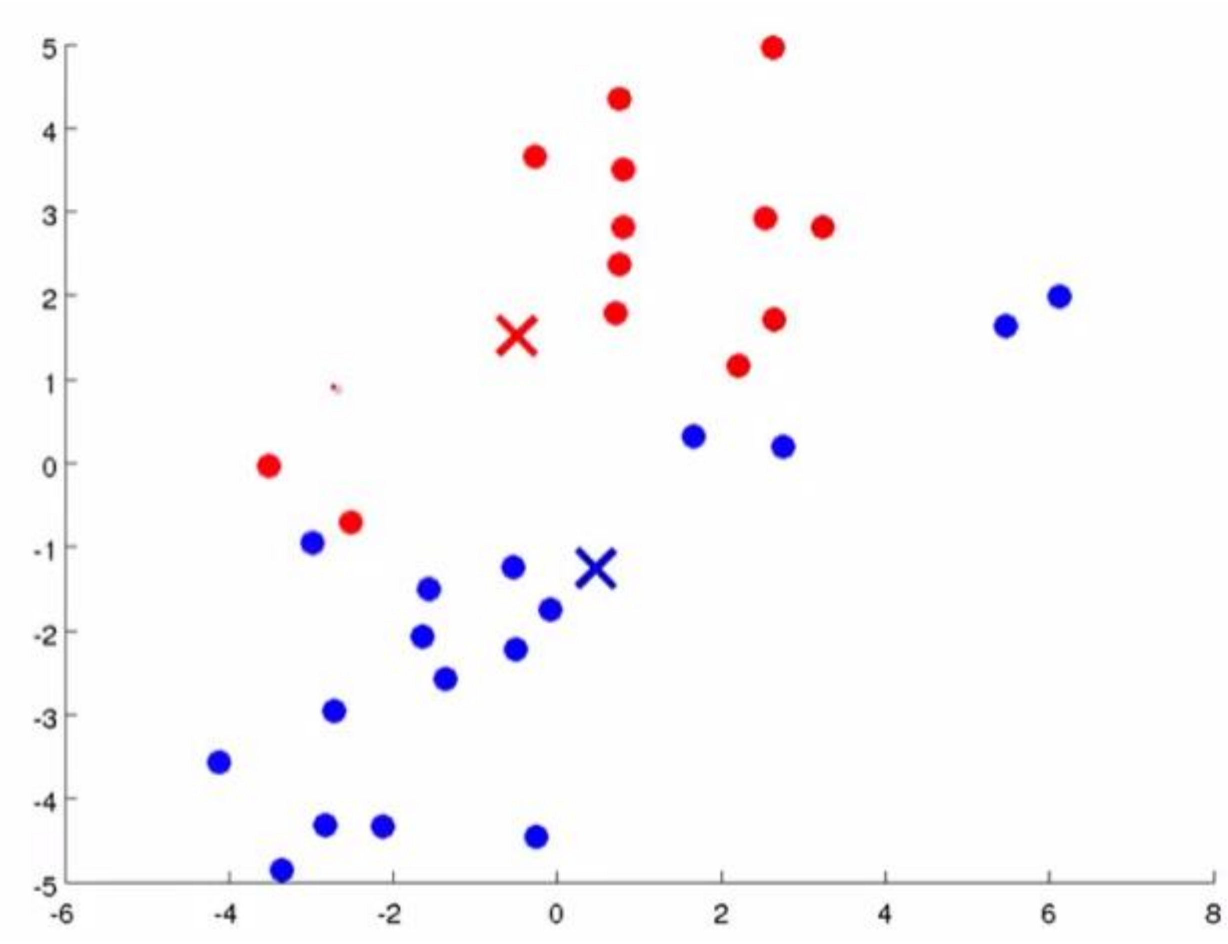
K-medias



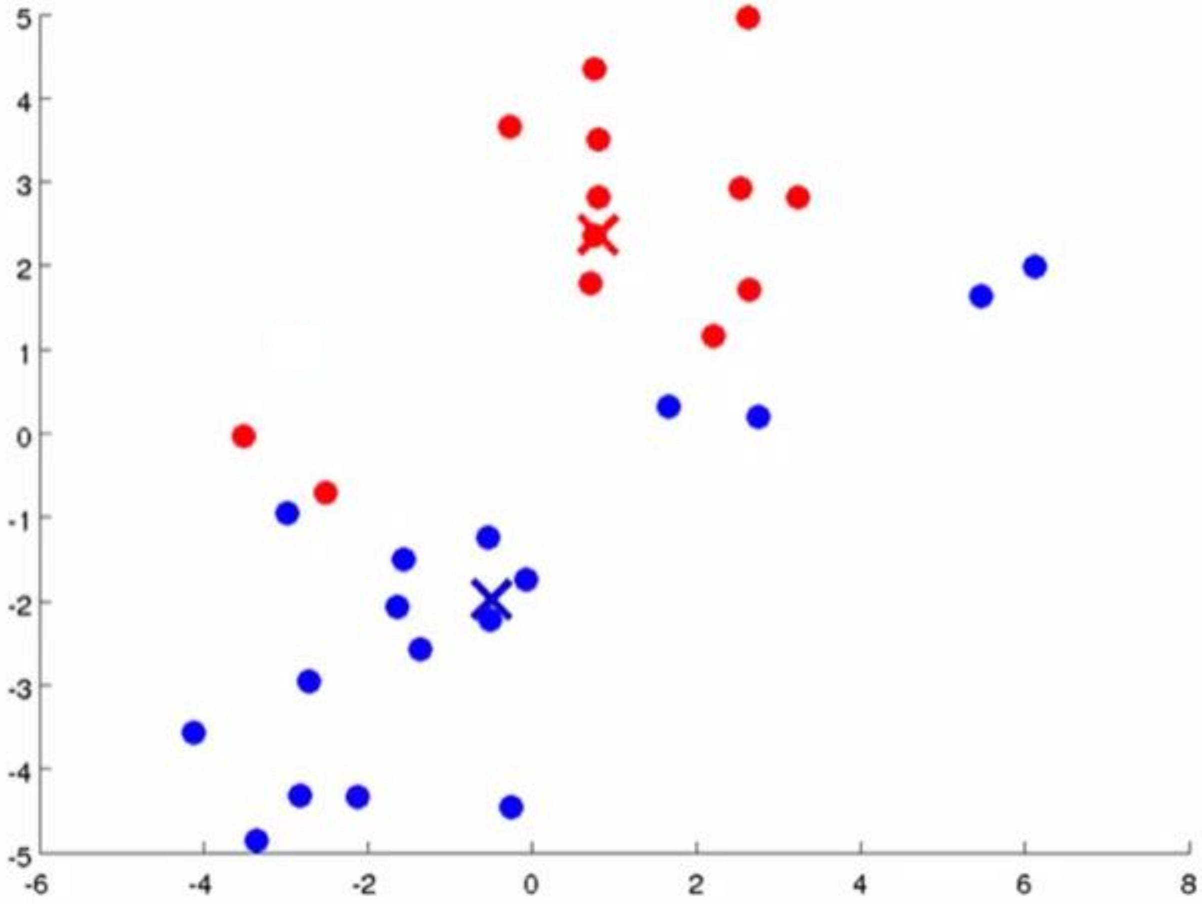
K-medias



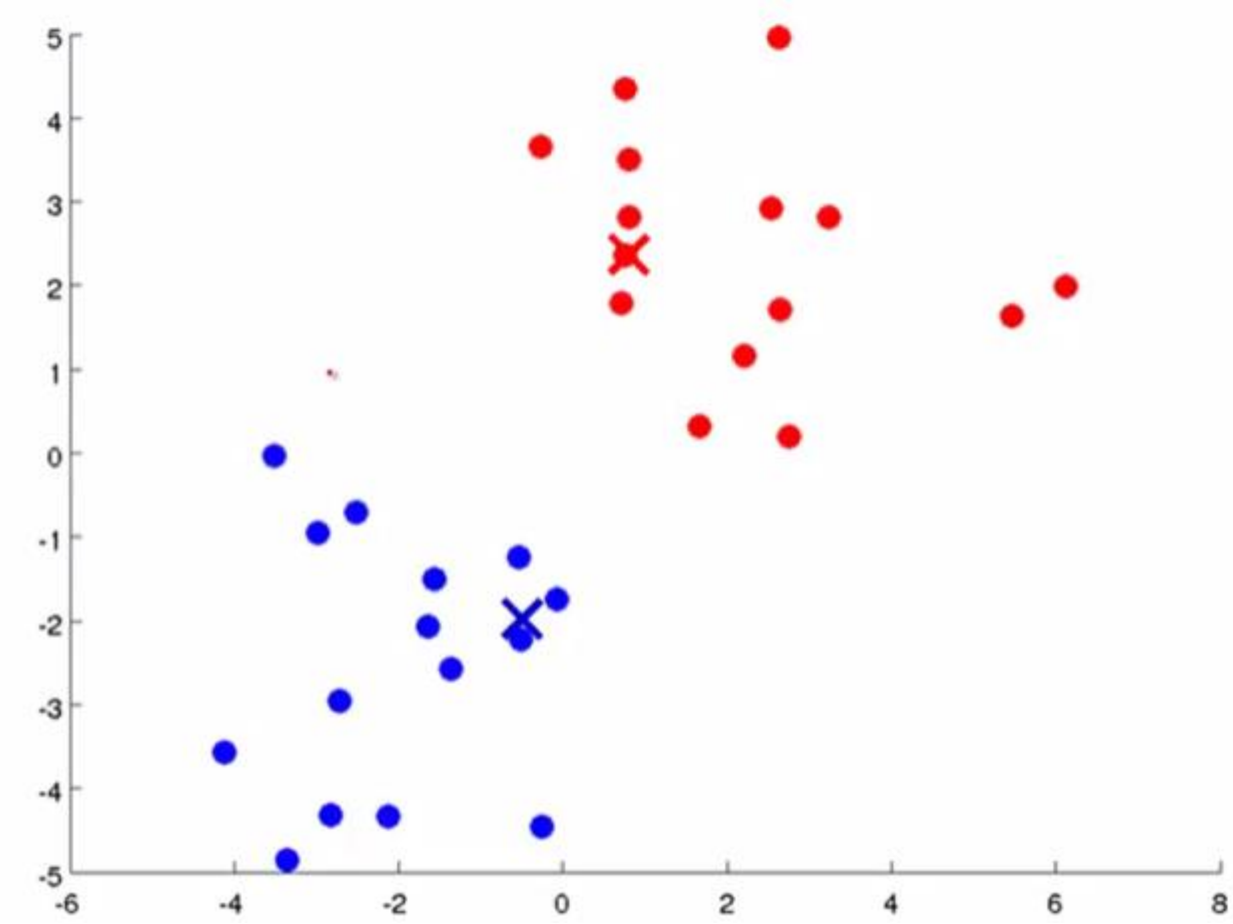
K-medias



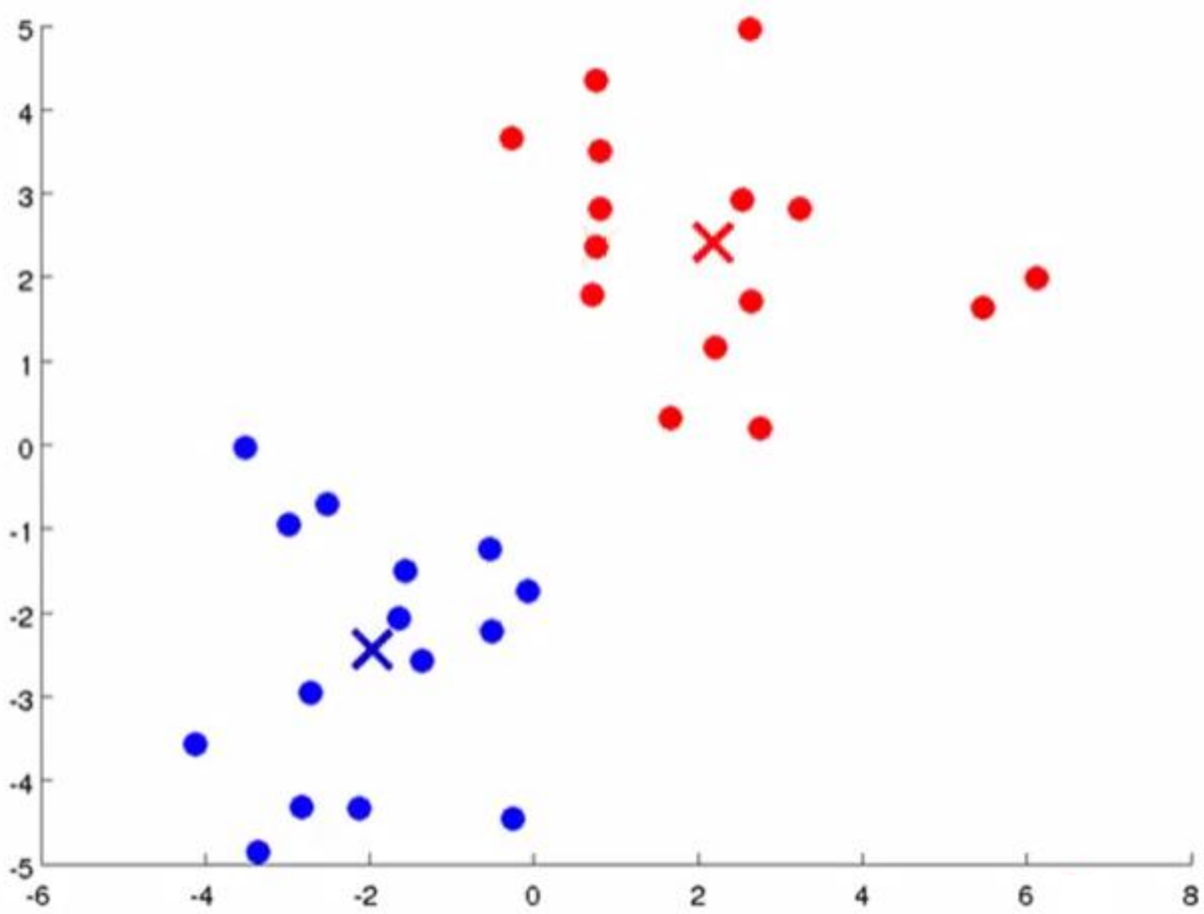
K-medias



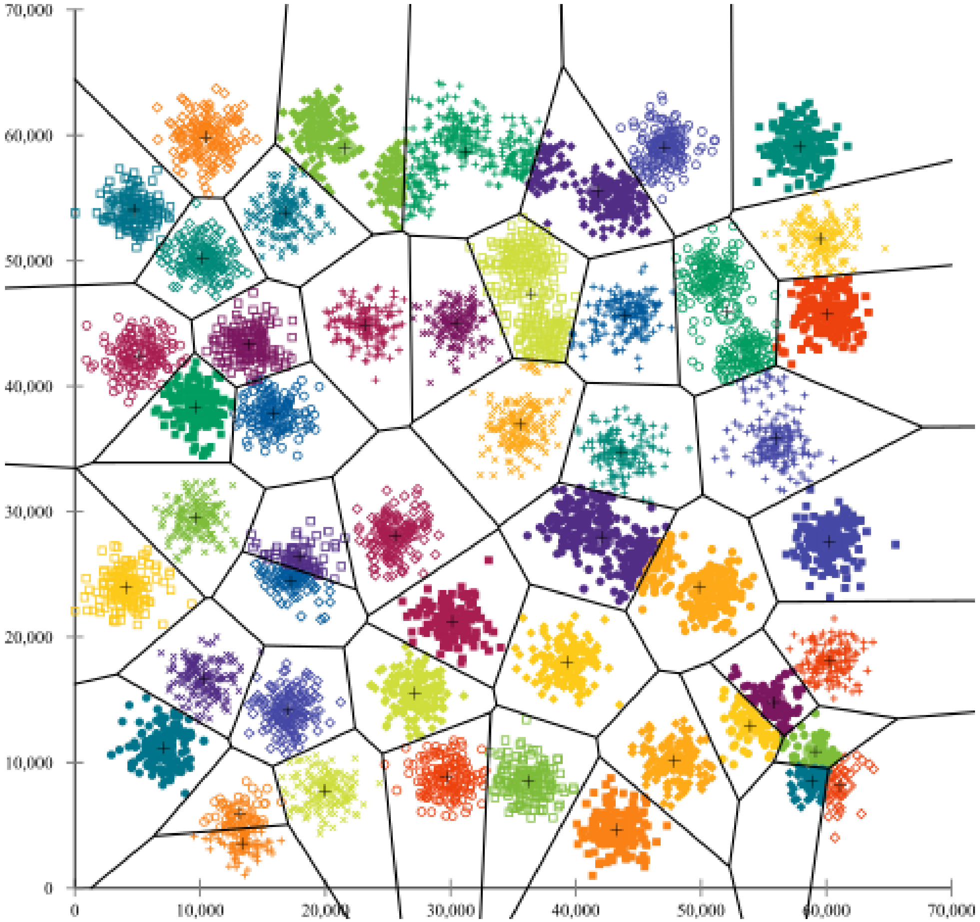
K-medias



K-medias



K-medias



Óptimos
locales

Publishing Web Services

Microsoft

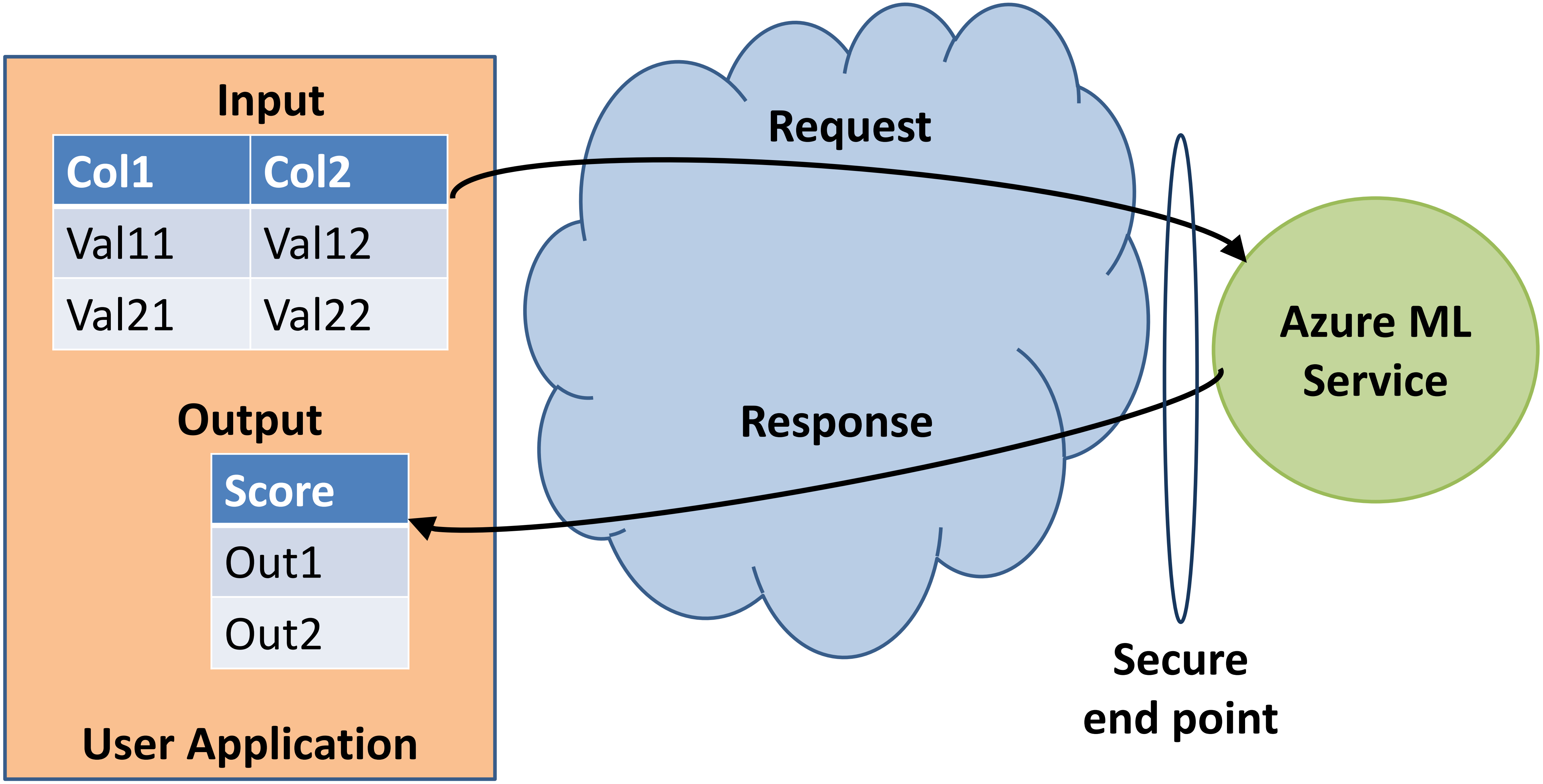
Overview

- Overview of web services
- Publishing models as a web service
- Using web services from Excel

Why Web Services?

- Publish machine learning solutions
  - Enable users to take action!
- Numerical and graphical results
- Integrate with desktop tools
  - BI tools
  - Excel
- Interactive and batch operations

What is a web service?



Azure ML Web Services Data Flow

Published Input

Input Transformations

Score Trained Model

Output Transformation

Published Output

Publishing Azure ML web services

- Automated process
- Creates end point
  - Unique URL
  - Keep API key secret!
  - RESTful API
  - Auto-generated code in C#, Python and R
- Secure HTTPS connection

Preparing experiment to publish

- Data must flow straight though
  - Process single row at a time
  - No aggregations!
- Most modules create a transform
- Remove any modules requiring multiple data values - no transform
- May need to edit R or Python code
- Add Select Columns module to end of data flow

Web Services Tips

- Reduce number of output columns
- Most modules create transforms
  - For example, scaling or model
- If no transform manual edit operations requiring multiple rows
  - Modules with no transform
  - Certain custom code

Retraining published web services

- Why retrain?
  - More data
  - Better model
- Simple update
  - Run Training Experiment
  - Update Scoring Experiment
  - Keep the schema the same!
- URL and key unchanged

Retraining published web services

- Why retrain?
  - More data
  - Better model
- Simple update
  - Run Training Experiment
  - Update Scoring Experiment
- URL and key unchanged

Azure ML Excel Web Services

- Downloadable Excel file
- Excel Azure ML plug-in
  - API key is secret
- Excel Online from OneDrive
  - Shareable
  - Batch or row at a time (RSS)

Microsoft