ANALYTIXLABS

Types of Business Problems

What are the Typical Business problems which we encounter?

What are the Typical Business problems?

- How to attract new customers?
- How to make those new customer to be profitable?
- How to avoid high risk/bad customers?
- How to understand the characteristics of current customers?
- How to make your unprofitable customers more profitable?
- How to retain your profitable customers?
- How to win back your existing customers?
- How to improve customer satisfaction?
- How to increase sales/profit and reduce expenses?
- How to recommend products to customers?
- How to optimize marketing expenses?
- How to take decision for offering credit card?
- How to increase credit line for given customer?
- How to optimize collection process?
- How to detect fraud transaction/customer?
- How to price the product?
- How to identify visitor will click or not?
- How to identify to employees who attrite?
- How to identify when customer stops buying/using product?
- How to predict how much customer make purchase?
- How to predict how much loss given the customer stop using product?
- how to calculate the impact of sales/volume given the price change?
- How to forecast the sales for next two quarters?
- How to optimize cash flows and funds utilization?
- How to optimize cash in ATMS?
- Does income of individual depend on demographics (Age and Years of education) and others?
- Which of the retail image levers drives footfalls or conversions?
- What drives satisfaction among branch users?
- What causes high performance of bank branch on the basis of financial parameters?

Lets deep dive some of the problems!

Example

In a credit card business. Applications have come for new card, bank has to take decision on whether to approve the credit or not and decide how much credit line need to be granted for given application?

Question

- Should we grant him/her the card?
- how much credit line need to be offered?

Non-deterministic information (Y)

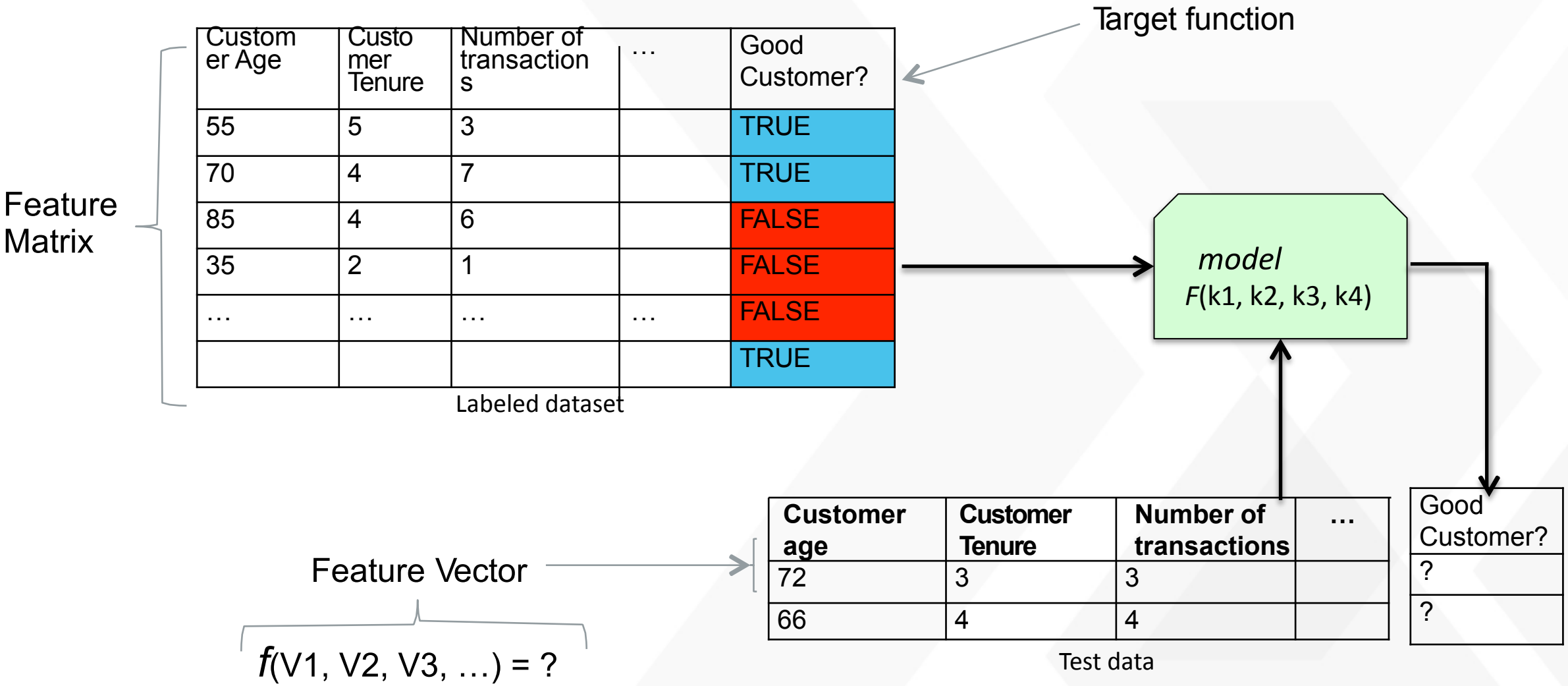
- Chances that the customer will default on his/her payments
- The maximum amount ($) that we may approve

Known information (X)

- Information on credit history, past transactions, financial status of the customer.

A Functional relationship between X and Y helps deciding whether to approve the credit request

Example- should we grant him credit card?



Are we predicting who is good/bad customer?

Example – How much credit line need to be offered?

Target function

Feature Matrix

Customer Age	Customer Tenure	Number of transactions	…	Credit Utilization
55	5	3		10000
70	4	7		8000
85	4	6		2000
35	2	1		20000
…	…	…	…	50000
				5000

Labeled dataset

model
F(k1, k2, k3, k4)

Feature Vector

f(V1, V2, V3, …) = ?

Customer age	**Customer Tenure**	**Number of transactions**	**…**
72	3	3	
66	4	4	

Test data

Credit utilization?
?
?

Are we predicting how much card utilization by given customer?

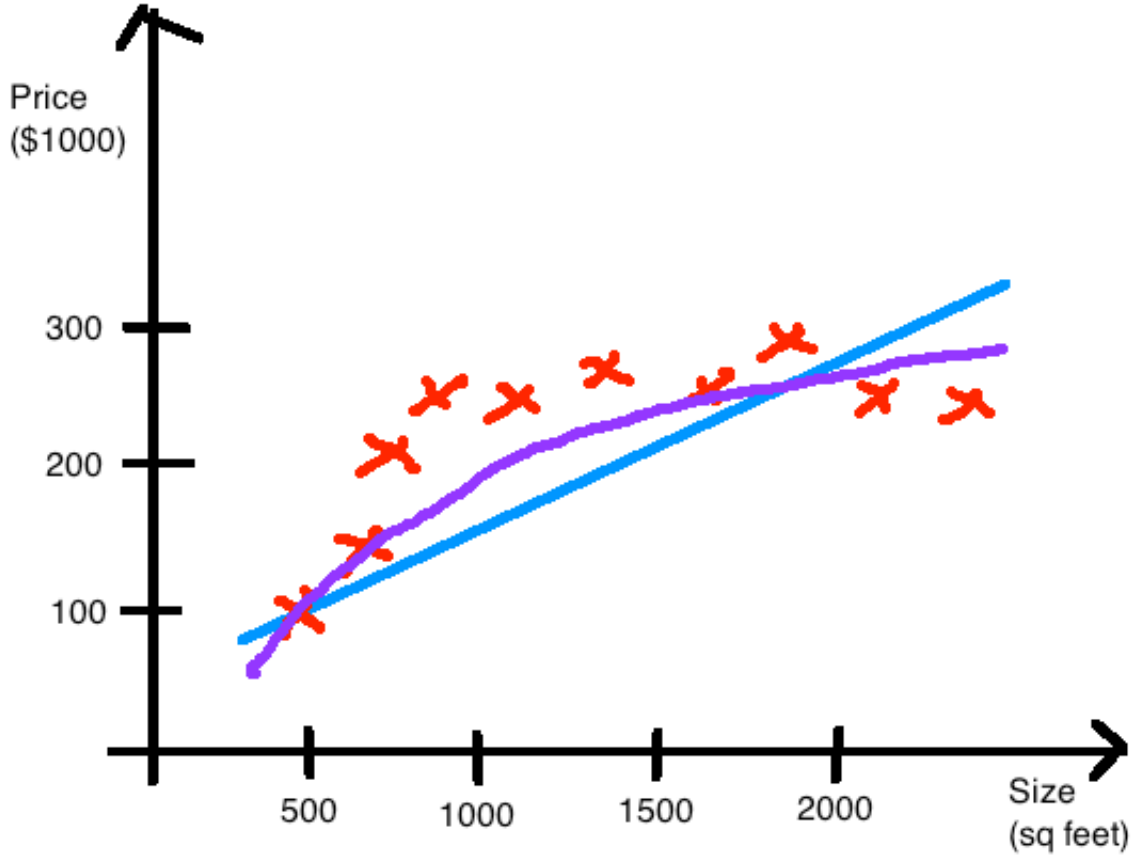
How to classify these problems?

Business problems – Types:

- Regression Problems – predicting a value
- Classification problems – predicting an event or predicting a category
- Segmentation problems – classifying the data when we don't have target variable(Un-supervised classification)
- Forecasting problems – Predicting future value(It is similar to regression however one of the independent variable is time)
- Others – rest of all like optimization problems, survival problems etc…

Regression Problems

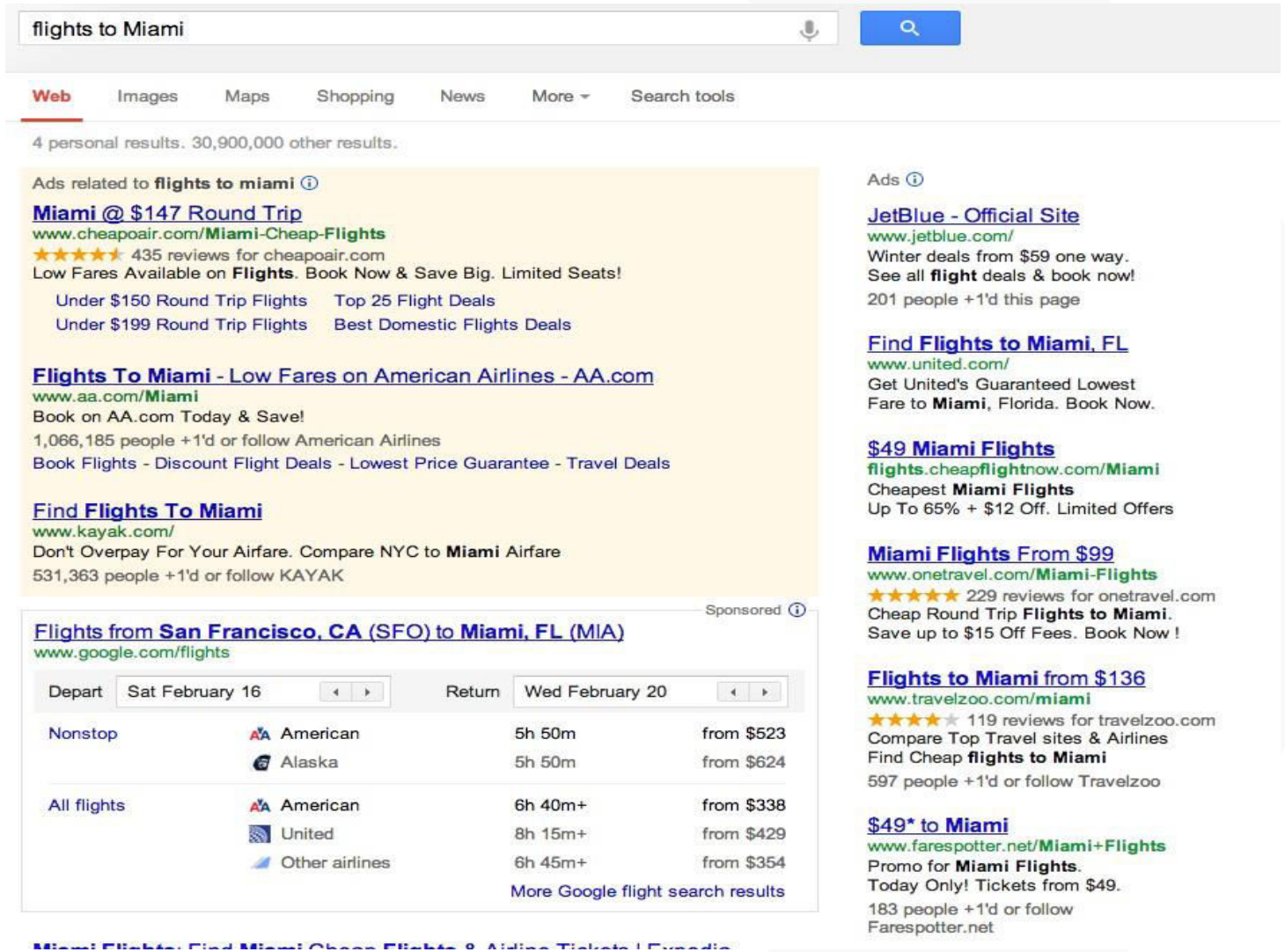
Regression: predict a continuous value



Some techniques:

- Linear Regression / GLM
- Decision Trees
- Support vector regression
- SGD
- Ensembles

Regression Example: Ad Click-Through Rates in Ad Search



Rank = bid * CTR

Predict CTR for each ad to determine placement, based on:

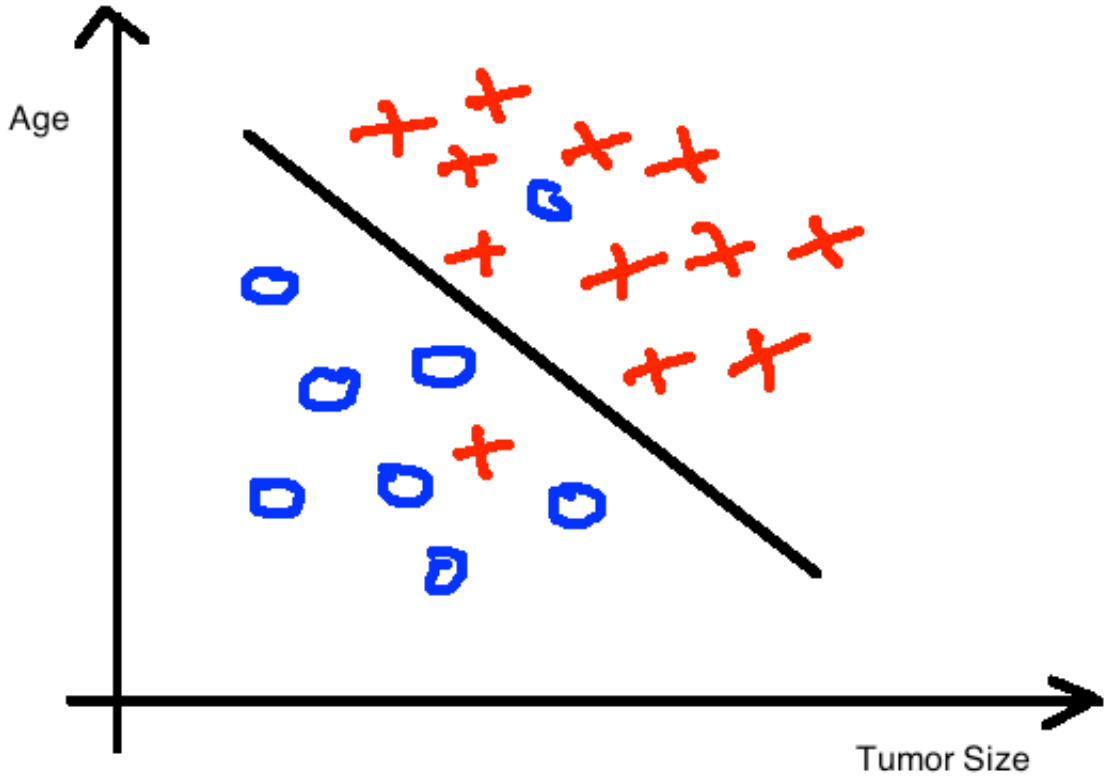
- Historical CTR
- Keyword match
- Etc…

Regression – Typical Applications

- Typical Applications:
  - Stock market: predict the share price for the future
  - Does income of individual depend on demographics (Age and Years of education) and others?
  - Which of the retail image levers drives footfalls or conversions?
  - What drives satisfaction among branch users?
  - What causes high performance of bank branch on the basis of financial parameters?
  - Energy demanding in a dam
  - Wind speed: eolic energy
  - Travel time prediction: for the planning of transport companies
  - Level of water in a river: for safety & prevention
  - Tax income: public budget
  - …

Classification Problems

Classification: predicting a category



Some techniques:

- Naïve Bayes
- Decision Tree
- Logistic Regression
- SGD
- Support Vector Machines
- Neural Network
- Ensembles

Detailed list of classification Techniques

Classical Techniques

- Logistic Regression
- Decision Trees (CHAID/CART)
- Linear Discriminant Analysis (LDA)
- Quadratic Discriminant Analysis(QDA)

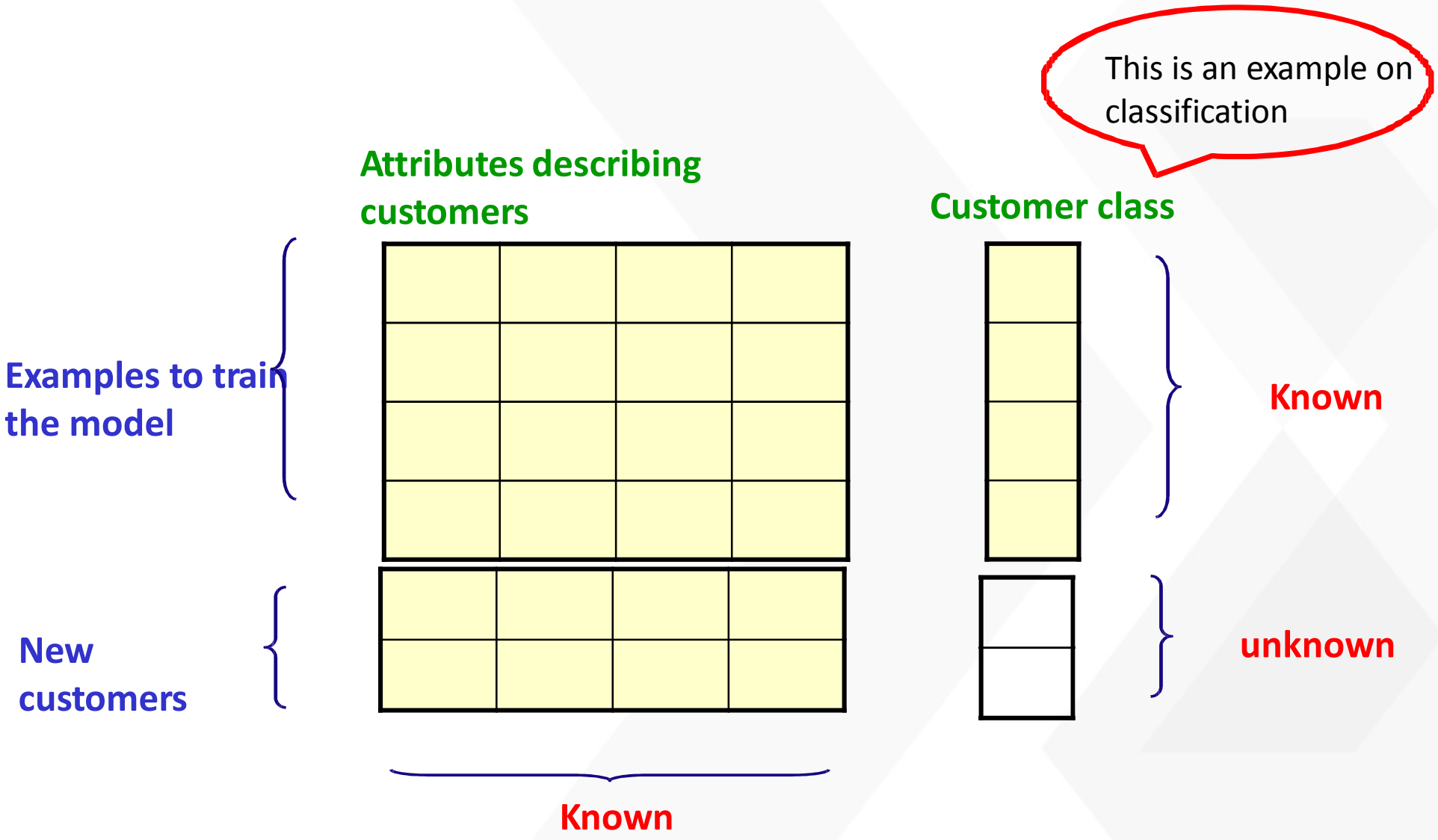
Ensemble Learning

- Bootstrapped Aggregation(Bagging)
- Boosting (AdaBoost/Gradient Boosting Machines)
- Random Forecast

Machine Learning Techniques

- K-Nearest Neighbours (KNN)
- Naïve Bayes
- Artificial Neural Networks (ANN)
- Support Vector Machines (SVM)

Classification Example

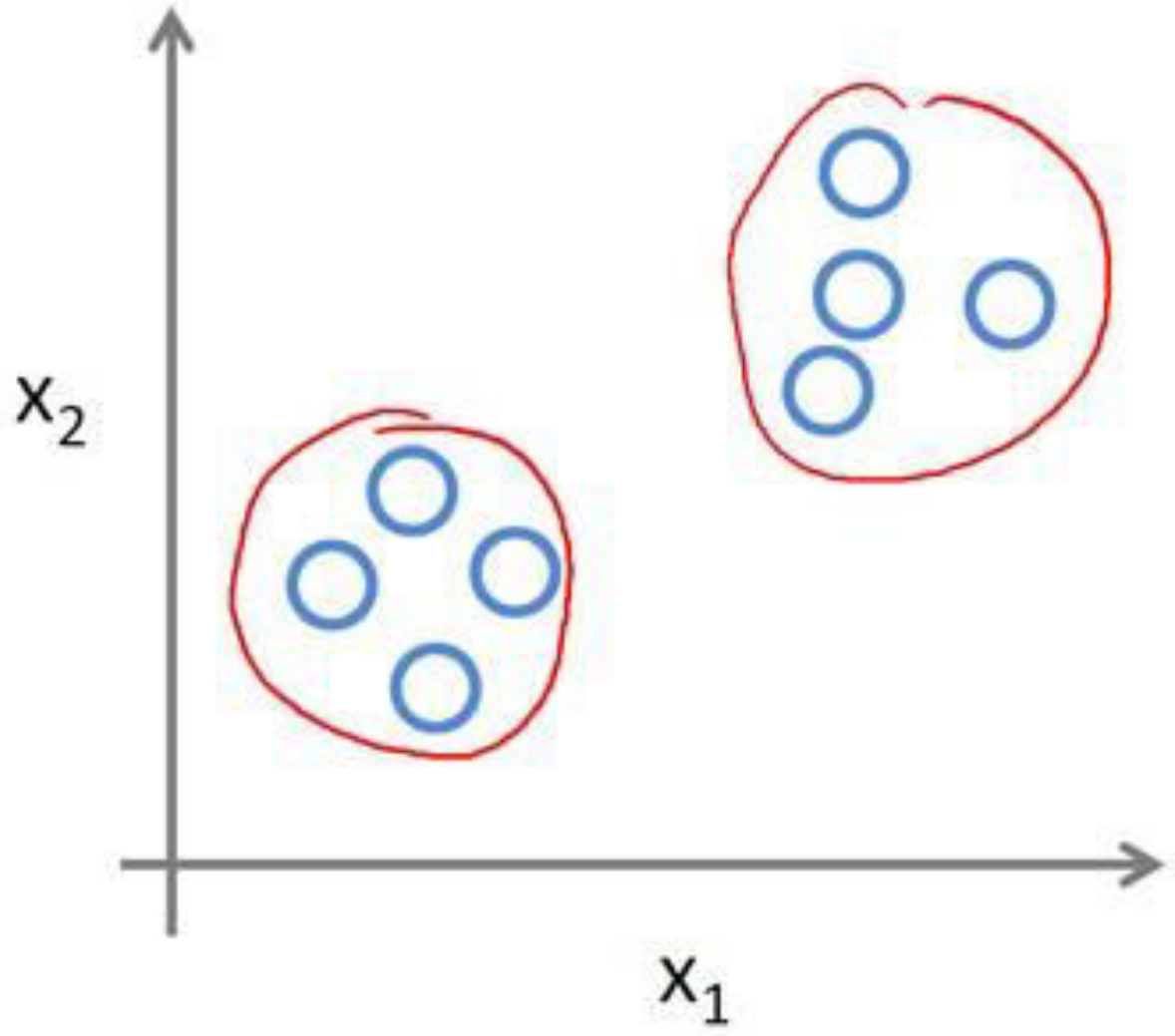


Classification – Typical applications

- Typical Applications:
  - Credit approval: classifies credit application as low risk, high risk, or average risk
  - Determine if a local access on a computer is legal or illegal
  - Target marketing (send or not a catalogue?)
  - Fraud Detection: Fraud Vs. Not Fraud
  - Collections: Identify cardholders that are likely to default and thus need collection effort (Payment Projection Models)
  - Insurance: Identify claims that are Fraud or Not Fraud
  - Marketing & Sales: Identify to responders to promotional campaigns(Response/Non Response, Buying/Not Buying
  - Operations: Models to identify to employees who attrite(Attrition/ Retention)
  - Website: Models to identify to weather visitor will click or not(Click/Not Click)
  - Gaming: Models to identify to who will win(Win/Loss)
  - Health Care: Models to identify to cure or not cure(Cure/ Not Cure)
  - Text classification (spam, not spam)
  - Text recognition (Optical character recognition)
  - …

Segmentation Problems

Segmentation: detect similar instance groupings / detect natural patterns

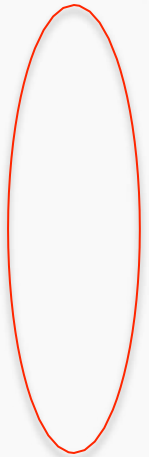


Some techniques:

- k-means
- Hierarchical clustering
- Spectral clustering
- DB-scan

Segmentation Example: Market Segmentation

Age	State	Annual Income	Marital status
25	CA	$80,000	M
45	NY	$150,000	D
55	WA	$100,500	M
18	TX	$85,000	S
…	…	…	…

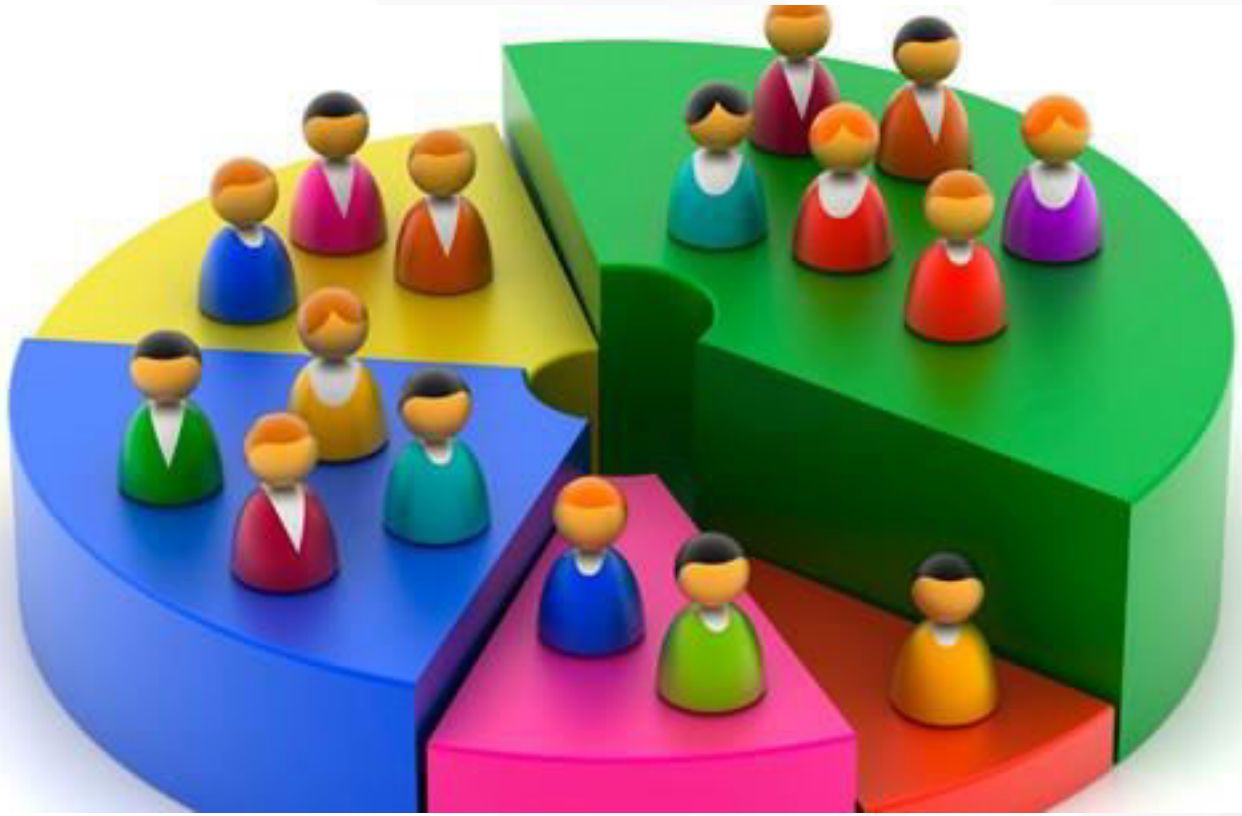


No labels

Model

Naturally occurring (hidden) structure

Example: market segmentation

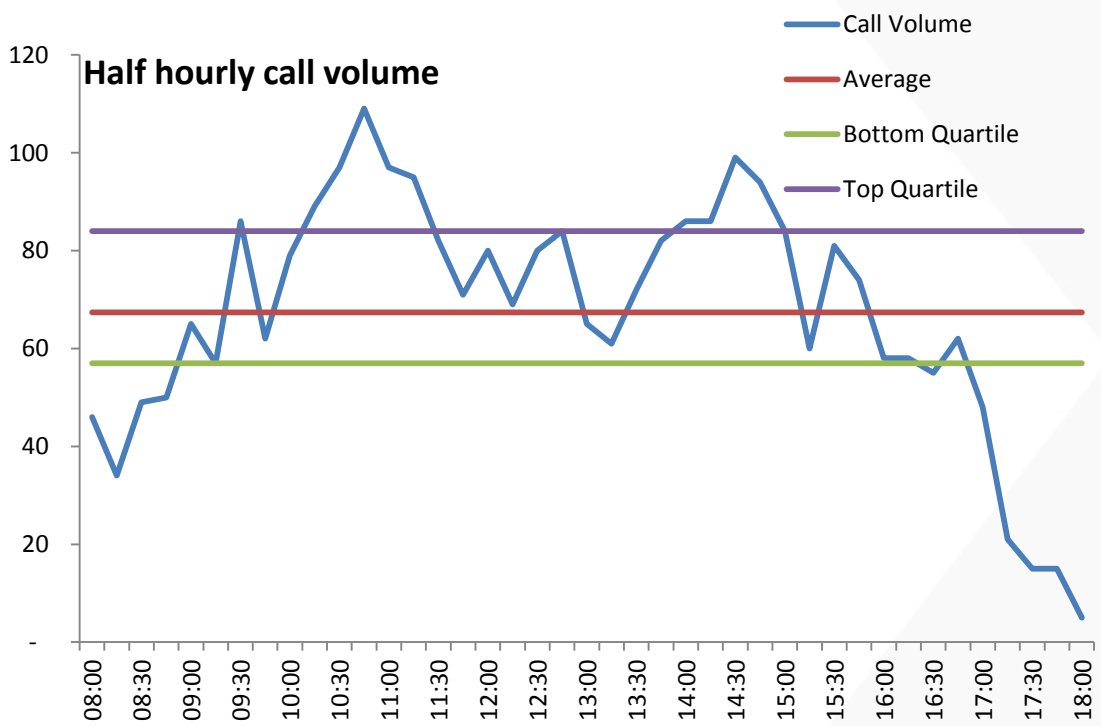


Segmentation – Typical applications

- Typical Applications:
  - Improve customer retention by providing products tailored for specific segments
  - Increase profits by leveraging disposable incomes and willingness to spend
  - Grow you business quicker by focusing marketing campaigns on segments with higher propensity to buy
  - Improve customer lifetime value by identifying purchasing patterns and targeting customers when they are in the market
  - Retain customers by appearing as relevant and responsive
  - Identify new product opportunities and improve the products you already have
  - Optimize operations by focusing on geographies, age groups etc. with the most value
  - Increase sales by offering free shipping to high frequency buyers
  - Offer improved customer support to VIP customers
  - Gain brand evangelists by incentivising them to comment, review or talk about your product with free gifts or discounts
  - Reactivate customers who have churned and no longer interact with you
  - …

Forecasting Problems

Forecasting: predict a continuous value for future(eg: next two quarters)



Some techniques:

- Averages
- Smoothening
- Decomposition
- ARIMA/SARIMA
- ARIMAX
- ARCH/GARCH
- VAR

etc…

Forecasting – Typical applications

- Typical Applications:
  - Call volume demand in call centers
  - Average handle time trends
  - Demand for seasonal maintenance
  - Event based demand for field services
  - Estimation of cash requirement in ATMs and Branches
  - Number of transactions for tellers
  - Footfall estimation in consumer retail
  - IT manpower requirement over months
  - …

Other Problems

Example: Affinity Analysis- identifying frequent item sets

	Item 1	Item 2	Item 3	Item 4	Item 5	…
Tx 1	Y	N	N	Y	N	
Tx 2	Y	N	N	Y	N	
Tx 3	Y	Y	N	Y	N	
Tx 4	N	N	Y	Y	Y	
Tx 5						
…						

	Item 1	Item 2	Item 3	Item 4	Item 5	…
Tx 1	Y	N	N	Y	N	
Tx 2	Y	N	N	Y	N	
Tx 3	Y	Y	N	Y	N	
Tx 4	N	N	Y	Y	Y	
Tx 5						
…						

Goal: identify frequent item set
Techniques: FP Growth, a priori

Example: Affinity Analysis

Use affinity analysis for
- store layout design
- Coupons

Predictive Modeling

What is Modeling?

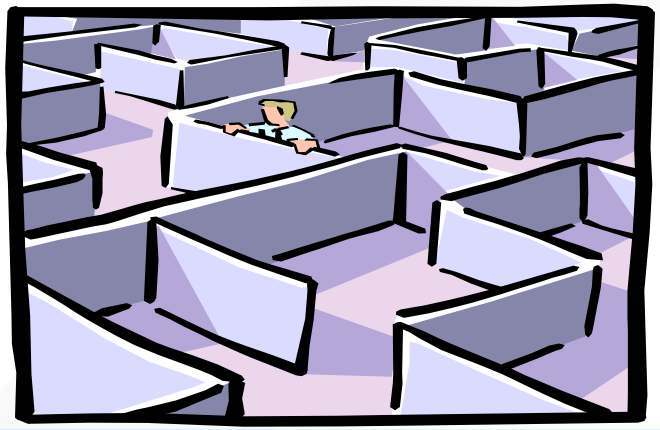
By “Modeling” we mean developing set of equations or mathematical formulation by which we can

- Predict certain events
- Identify the drivers of certain events based on some explanatory variables

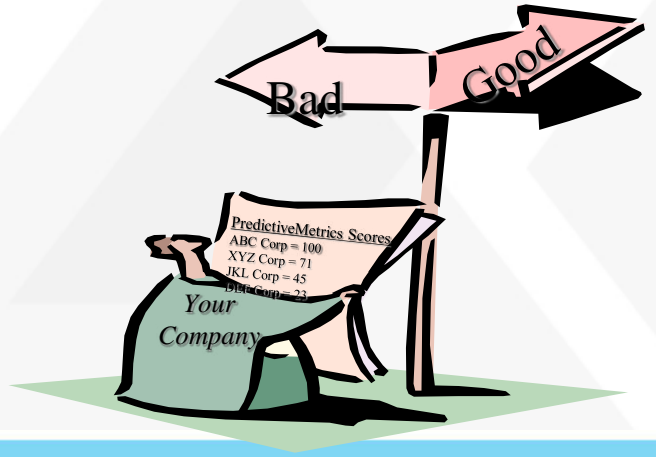
For example, we can build models to predict drivers of sales, risk of a borrower.

Historical Data

Statistical Analyses



Predict Future Events



What is predictive Modelling?

Predictive Model:

Goal: To predict the value of a given variable (named **target** or **objective** variable)

$$y = f(\mathbf{x})$$

- Training: given a training set of labeled examples $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$, estimate the prediction function f by minimizing the prediction error on the training set
- Testing: apply f to a never before seen test example x and output the predicted value **y = f(x)**

- For each record on the dataset determines the value of the class attribute
- Constructs a model based on the training set; then, uses the model in predicting new data

Why Do We Estimate f?

Predictive Modeling is all about how to estimate f.

Why do we care about estimating f?

There are 2 reasons for estimating f,

- ✓Prediction
- ✓Inference

Predictive Models: Examples

- Prospecting/Response model (stage 1): predict potential customers' likelihood of conversion
- Xsell model (stage 1): predict current customers' likelihood to purchase other products
- Balance model (stage 2): predict customers' opening balance if they open accounts
- Potential value model: measure customers' future (profit) potential
- Risk/Credit model: predict people's likelihood of default/charge off

Why do we need predictive Model?

- Distinguish/understand different types of customers in term of risk, potential value, likelihood of conversion/xsell/attrition, etc.
- Model helps us better targeting audience – higher conversion rate, lower marketing cost
- Compare with human judgment, model costs less and is more consistent, robust, efficient – easy to implement on large population
- With new information, model can be systematically evaluated and improved to enhance targeting

Overview of modeling key concept

Term	Examples	Description
Target / Y /Dependent Variable	Paid users among all the users	Represents the output or effect
X / Independent Variable	Age, gender, product usage	Represent the inputs or causes
Probability	26% to become paid user	Is a measure of the expectation that an event will occur or a statement is true
Score	0.26	Is the value of a variable below which a certain percent of observations fall
Percentile	75%	Is the value of a variable below which a certain percent of observations fall

Define target variables

- **Look-a-like Model:** Use customer who are currently having the product as modeling target
- **Walk-in Model**: There is a modeling window. Customer who opened the product during this time period is defined as modeling target
- **Response Model**: Use customer who converted as a result of campaign as modeling target
- **Uplift Model**: Use the change in behavior as a result of a treatment as modeling target

	Pros	Cons
Look-a-like	When there are not enough modeling targets, look-a-like model is best way to remedy sample size issue.	Model works like a profile. It uses the differences between product holder and non-holder as main drivers. It could be misleading in cause-effect and event sequencing
Walk-in	There is a time window. Model captures the natural response. It's a good start when no campaign was ever launched.	It's still a retrospective model, not campaign driven. Does not capture marketing effect
Response	Uses the results of real campaigns. Natural response + marketing effect	Smaller sample size. Non-represetative sample of population - cut based on BAU and old models
Uplift	Identify the pursuable that will actually be influenced by your campaign, avoid targeting individuals that will buy anyway	No proven techniques yet to achieve reliable results as other modeling type

Nature of Explanatory & Dependent variables

An Explanatory variable could be

- ✓ Numerical
  - Discrete : e.g. Number of satisfactory trades
  - Continuous : e.g. Highest Credit Line
- ✓ Categorical
  - Ordinal : e.g. Income Group (High/Medium/Low)
  - Nominal : e.g. Gender (Male/Female)

A Dependent variable could be

- ✓ Continuous: e.g. The total ($) that we may approve
- ✓ Discrete : e.g. Number of equipments that may be funded
- ✓ Binary : e.g. Whether the customer would default on payment or not (1/0)

Analyze Data Major Steps

Steps	
0 Business Problem	▪ Convert business problem into statistical problem ▪ Identify type of problem - Technique ▪ Define Hypothetical relationship
1 Data Construction	▪ Create the model data by various sampling ▪ Aggregate the data at same level(eg: customer level) – Depends technique
2 Univariate/Bi-Variate Analysis	▪ Examine the data for errors, outliers and missing values. ▪ Assess/understand the relationship to target variable ▪ Understand the relationship between independent variables
3 Data Preparation	▪ Exclusions/Data type conversions/Outlier treatment/Missing value treatment ▪ Create new, hypothetically relevant variables, e.g. max, min, sum, change, ratio ▪ Binning variables – dummy variables creation ▪ Transform data to help ensure linearity
4 Variable Reduction	▪ Avoid collinearity and shorten computing time by reducing the number of independent variables – variable cluster, correlation, factor analysis etc.

Variable reduction techniques

The following variable reduction techniques have been using as part of model development.

- ✓ Information Value or Weight-of-Evidence
- ✓ Principal Component Analysis /Factor Analysis
- ✓ Variable Classing (Variable clustering)
- ✓ Variance Inflation Factor(VIF) / Conditional Index(CI)
- ✓ Marginal Information Value
- ✓ Step-wise Variable Selection (Forward/Backward/stepwise)
- ✓ Univariate Analysis

Model Development Major Steps

	Steps
1 Split the data for validation	▪ Split the file into the modeling(training) and validation(test) data sets
2 Build the Model on Training data	▪ Linear Regression/Logistic Regression/Decision Trees/Segmentation etc. ▪ Assumptions checks ▪ Modify the data as per the assumptions(if required)
3 Process the model	▪ Interpret the model ▪ Understand model diagnostics – accuracy(fit) of the model ▪ Iterate the models
4 Validate the model	▪ Re-run the model ▪ Using scoring ▪ Using cross validation(K-Fold) Validation
5 Implementation of the Model	▪ Prepare final model results– present the model ▪ Identify the limitations of the model ▪ Implement the model(converting stats solution into business solution)

Regression Modeling

Business Problem

I am the CEO of a hypermarket chain “Safegroceries” and I want to open new store which should give me the best sales . I am hiring “Alabs” to help me figure out a location where to open the new store

What should ALABS do ?

Additional Information about Safe groceries:

- Safegroceries has more than 5000 stores across the world
- It is upstream hypermarket store catering to high end products
- There are more than 100 locations he needs to choose from ?

What could impact sales ?

- ✓ Population Density in the area
- ✓ Disposable Income
- ✓ Demographics of the region
- ✓ Parking size of the location
- ✓ No of other grocery stores in around (3km)
- ✓ Credit card usage
- ✓ Internet penetration/usage
- ✓ Average no of cars/household
- ✓ Avg family size/household
- ✓ No of working people/household
- ✓ ……..
- ✓ ……..

Relationship between Sales and Variables

- ✓ Sales = function (X1, X2, X3, X4,X5,X6……)
- ✓ Sales = 10X1 + 20X2 +0.5X3 + 8X4 +……
- ✓ If the function is linear we call it linear regression

This was a case of prediction . How about doing root cause analysis ?

Now CEO wants to improve the performance of the existing stores and wants to increase sales ?

Decision – Prediction vs Inference(root causal)

Regression

Regression Analysis

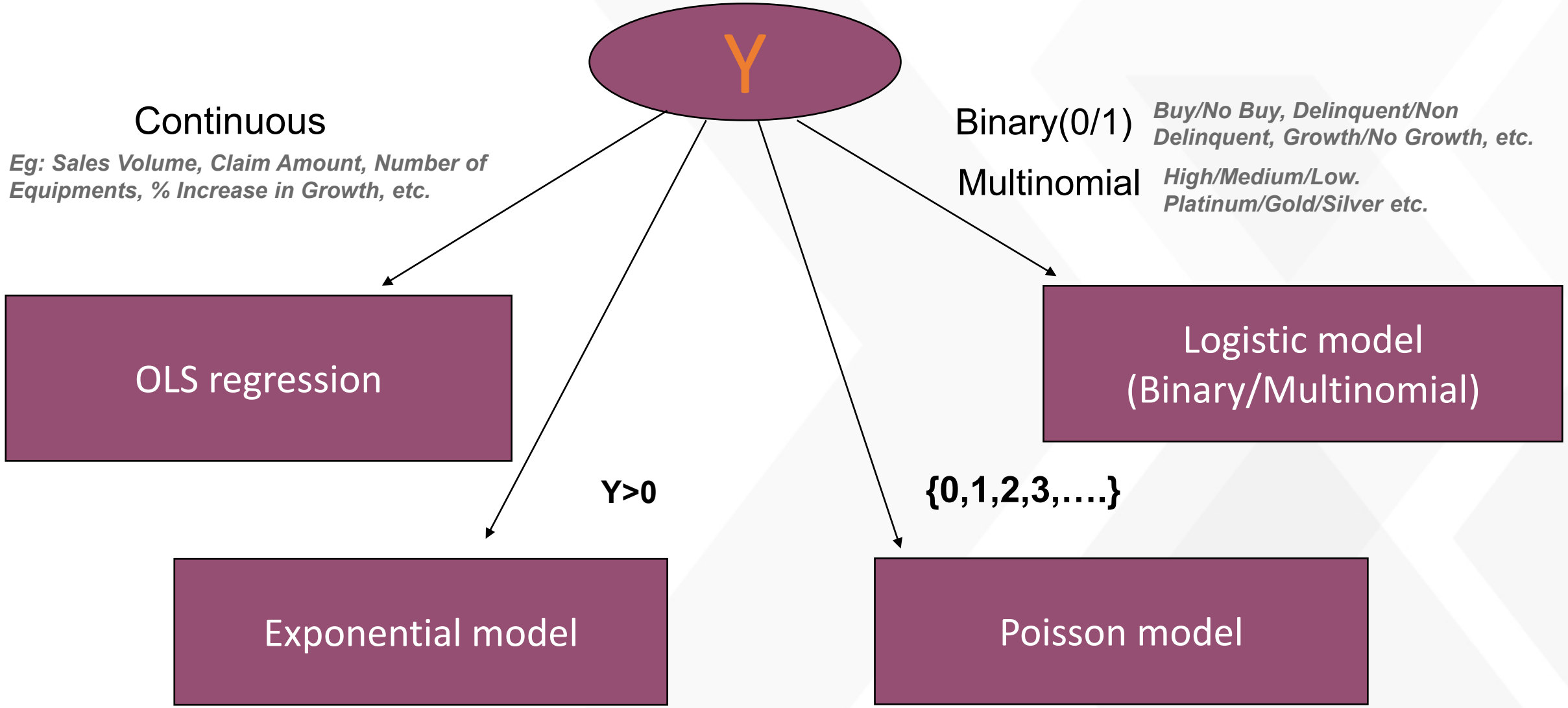
"Regression analysis is a statistical tool for the investigation of relationships between variables. Usually, the investigator seeks to ascertain the causal effect of one variable upon another"

Regression modeling

Establishing a functional relationship between a set of Explanatory or Independent variables X_1, X_2, …, X_p with the Response or Dependent variable Y.

$$Y = f(X_1, X_2, .., X_p)$$

Types of Regression Models

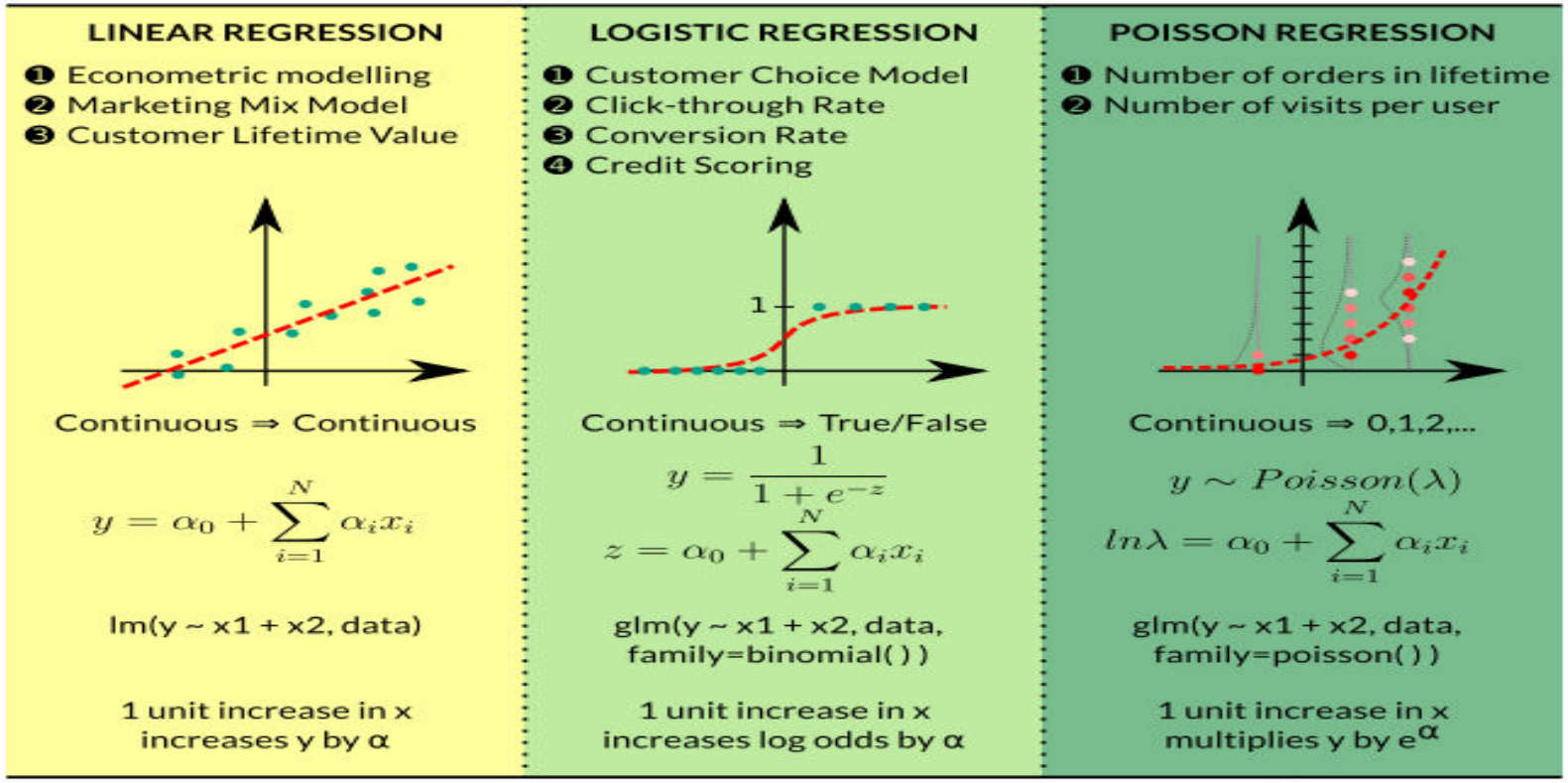


Three Regression Types (GLM)

Generalized linear models extend the ordinary linear regression and allow the response variable y to have an error distribution other than the normal distribution.

GLMs are:

- A. Easy to understand
- B. Simple to fit and interpret in any statistical package
- C. Sufficient in a lot of practical applications



Contact us

Visit us on: http://www.analytixlabs.in/

For course registration, please visit: http://www.analytixlabs.co.in/course-registration/

For more information, please contact us: http://www.analytixlabs.co.in/contact-us/

Or email: info@analytixlabs.co.in

Call us we would love to speak with you: (+91) 88021-73069

Join us on:

Twitter - http://twitter.com/#!/AnalytixLabs

Facebook - http://www.facebook.com/analytixlabs

LinkedIn - http://www.linkedin.com/in/analytixlabs

Blog - http://www.analytixlabs.co.in/category/blog/