

The slide features abstract green geometric shapes. On the left is a tall, thin green triangle pointing downwards. On the right is a larger, more complex shape composed of several overlapping green triangles and polygons in various shades of green. A thin grey line extends from the bottom left towards the right, passing behind the green shapes.

The Data Incubator Scholar program interview

Amin Shafaat

About myself

- ▶ Name
 - ▶ Mohammad Amin Shafaat
- ▶ Education
 - ▶ PhD Candidate-Engineering Sciences and Mechanics (2015)
 - ▶ Pennsylvania State University-State College PA
 - ▶ Master of Science-Materials Science and engineering (2009)
 - ▶ Amirkabir University of Technology, Tehran-Iran
- ▶ TDI Capstone Project
 - ▶ Model for predicting the date in which a product is most likely to be bought for an Online Retail company
- ▶ PhD Research Project
 - ▶ Machine learning approach in particle sensing using Whispering Gallery Mode Resonator

Capstone Project

- ▶ Title: Model for predicting the date in which a product is most likely to be bought for an Online Retail company
- ▶ Data set: Online Retail data set which contains all the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail for 541909 observations.
- ▶ Products: unique all-occasion gifts
- ▶ Variables
 - ▶ InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, Country

Importance of the study

- ▶ Multivariate logistic regression would help us determine the date in which a specific product is most likely to be bought.
- ▶ Also, by grouping the gifts, sales department would be able to offer specific group of customers a group of products that they are likely to buy.
- ▶ Similar analysis can be performed for cancelled items for the company itself and prevent waste of products and money.

Analysis

► Cluster Analysis

Cluster	Description
1	Buy only In February 77184\$/month
2	8 months of interaction 299\$/month
3	2 months interaction 367\$/month
4	12 Month of interaction 709\$/month
5	Last 4 months of year 14057\$/month

► Multivariate logistic regression

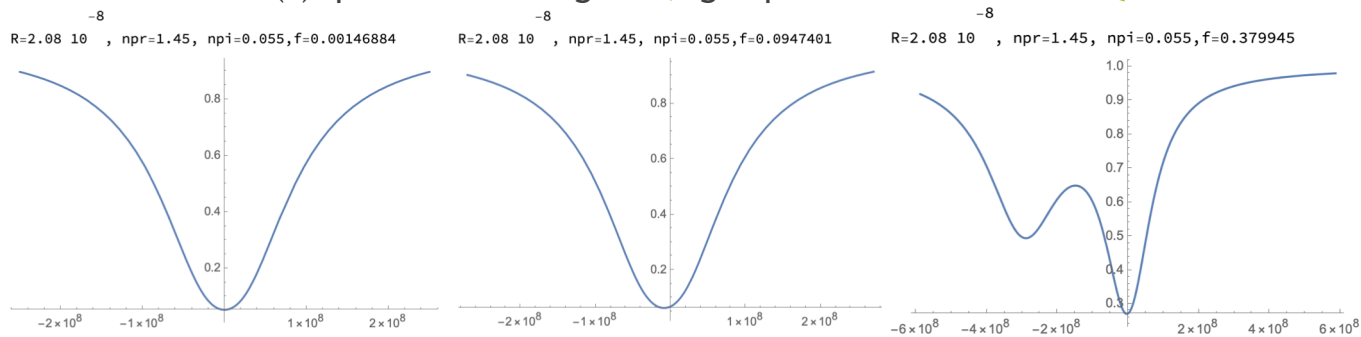
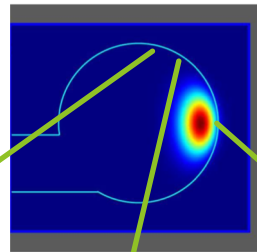
- After data cleaning I try to make contingency table but presently I face memory problem with R

```
Error in table(x, exclude = exclude) :  
  attempt to make a table with >= 2^31 elements
```

Derivation of transmission curve

Comments about the random parameter $f(r)$

- ▶ Intensity of light varies inside the resonator.
- ▶ The higher the intensity, the stronger the interaction of particle with light.
- ▶ the particle hits the resonator randomly.
- ▶ Random number $f(r)$ quantifies strength of light-particle interaction.



Procedures

1. Mathematica code was developed to simulate transmission formula at different particle parameters:
 - ▶ Properties of particles:
 - ▶ Size (R): 1-95 nm with 5 nm interval
 - ▶ Real part of the refractive index (np_r): 1-2 with 0.05 interval
 - ▶ 100 repetitions for each (R, np_r) combination
 - ▶ Computed parameters:
 - ▶ Full transmission curves
 2. Finding Principal Components for transmission curves
 3. Applying Discriminant analysis to classify them based on R and np_r
- ▶ Results: Discriminant Analysis can classify transmission curves with overall error rate 10% for particle sizes ranging 20-70 nm.

Thank you for your attention