

# Predicting Conversions: Click to Buy

Sarthak Jindal

Indraprastha Institute of Information Technology

Email: sarthak15169@iiitd.ac.in

## 1. Motivation

The information of the probability of conversion of a click to a buy on an E-Commerce website can be used to intelligently display items that are more probable of being bought.

## 2. Problem Statement

Given for each click:

```
"item_price_level",  
"item_sales_level",  
"item_collected_level",  
"item_pv_level",  
"user_gender_id",  
"user_age_level",  
"user_star_level",  
"context_page_id",  
"shop_review_num_level",  
"shop_review_positive_rate",  
"shop_star_level",  
"shop_score_service",  
"shop_score_delivery",  
"shop_score_description"
```

To predict whether the click converted to a buy or not.

## 3. Dataset

Taken from IJCAI-18 CVR Prediction Contest.

Total number of data points = 277000

Dimension of each data point = 14

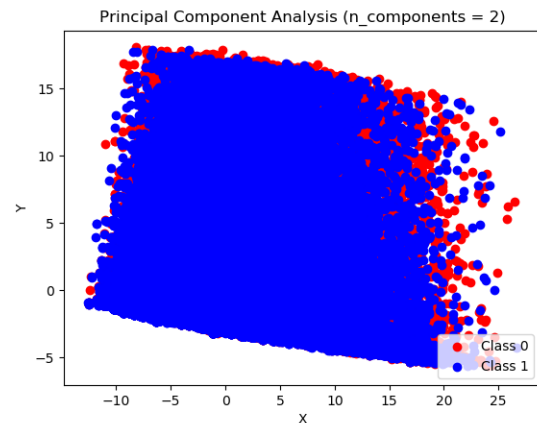
Class Prior Ratio = 1:1

## 4. Data Visualization

### 4.1. Principal Component Analysis

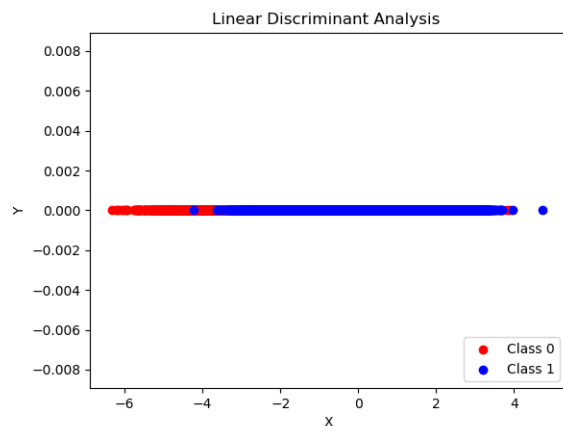
Clearly, the projected data overlaps. Consequently, the classification accuracy using PCA

will be low as seen in the following section.



### 4.2. Linear Discriminant Analysis

Clearly, the projected data is separable. Consequently, the classification accuracy using LDA will be high as seen in the following section.

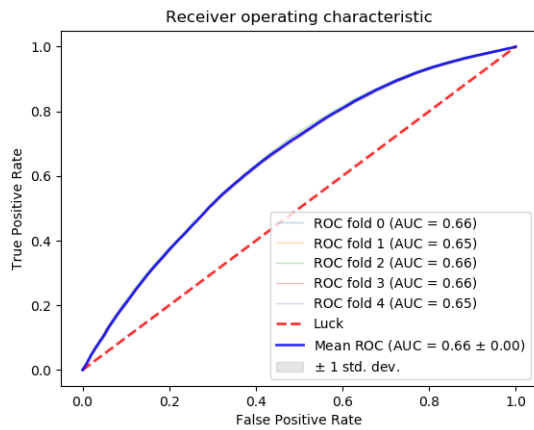


## 5. Feature Selection

Filter: For a given K, K- best features selected according to least mutual dependence.

Wrapper: For each K from 1 to 14, obtained accuracy on K- best features.

Result: 10- best features gave highest classification accuracy.

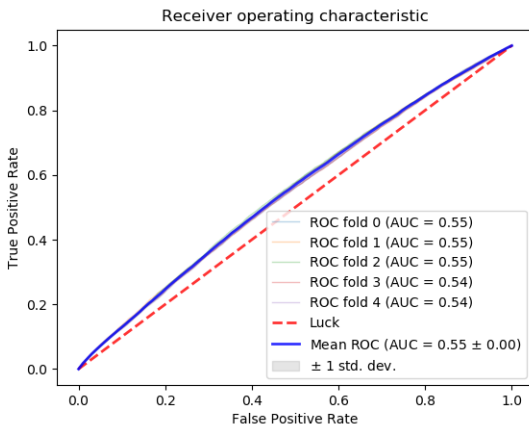


## 7. Analysis

As is evident from the visualisations and ROCs, the data was best classified after feature extraction by LDA. Different classifiers used- Naive Bayes, Neural Networks, AdaBoost used- all give similar results due to simple data distribution.

## 6. Feature Extraction

### 6.1. Principal Component Analysis



### 6.2. Linear Discriminant Analysis

