

MTH 443: Lab Problem Set 2

Dataset #1: “World economic development” dataset (source: International Monetary Fund) (eco_dev_data.csv) containing data on the following economic development indicators of 121 countries.

Economic Development Indicators

Indicator and abbreviation	Aspect of Economic development
GNP per capita at PPP (GNPPER)	Income level
GDP growth rate (GDPGR)	Growth of economy
Gross domestic investment as percentage of GDP (DOMINV)	Level of investment
GDP deflator (GDPDFL)	Inflation
Agriculture value added as percentage of GDP (AGRVLAD)	Structure of output
Industry value added as percentage of GDP (INDVLAD)	Structure of output
Export of goods and services as percentage of GDP (EXP)	Openness of economy
General government consumption as percentage of GDP (GOVCON)	Role of government
Resource balance as percentage of GDP (RESBL)	Net borrowing/lending on account of merchandise trade
Domestic credit provided by the banking sector as percentage of GDP (DOMCRDT)	Private sector financing
Ratio of gross international reserve to imports (GRIIMP)	Strength of foreign exchange reserve
Number of months of import cover (IMPCOV)	Strength of foreign exchange reserve
Interest spread (INTSPRD)	Efficiency of financial market

Dataset #2: “Car dataset” (cars_data.csv) gives data of 32 cars on 11 variables which comprise of important features of the automobile. Variable codes given in the dataset and their description is given below:

Variable code	Variable description
Mpg	Miles/(US) gallon
Cyl	Number of cylinders
Disp	Displacement (cu.in.)
Hp	Gross horsepower
drat	Rear axle ratio
wt	Weight (1000 lbs)
qsec	1/4 mile time
vs	Engine (0 = V-shaped, 1 = straight)
am	Transmission (0 = automatic, 1 = manual)
gear	Number of forward gears

Dataset #3: “eur_protein_consump.csv” dataset gives protein intake of European countries. Intake level of the 9 major proteins in 25 countries are reported: (1) red meat, (2) WhiteMeat, (3) eggs, (4) milk, (5) fish, (6) cereals, (7) starch, (8) nuts and (9) fr. Veg.

Dataset #4: “Bank financial ratios data set’ (Source: Reserve Bank of India) (PS_bank_fin_ratio.csv) containing important financial ratios of Indian public sector banks during the period from financial year 1996-1997 to financial year 1999-2000.

Ratio	Abbreviation
Return on Equity	ROE
Return on asset	ROA
Cost of deposit	COD
Cost of borrowing	COBR
Return on advances	ROAD
Return on Investment	ROI
Operating profit to total asset	OPTAST
Net Interest income to total asset	NIITAST
Spread=Return on (advance+investment)-cost of deposits	SPRD
Staff expenses to total expenses	STEXTEx
Net NPA to net advances	NET_NPA
Capital adequacy ratio	CAR
Business per employee	BUSEMP
Profit per employee	PFTEMP

For each of the datasets

- (a) Obtain principal component projection. Keep the case tag (identifier) on the projected plane so as to know location of the cases on the plane.**
- (b) Find the proportion of total variation explained by the principal components**
- (c) Obtain scree plot and observe the elbow formation and hence suggest the number of principal components to retain for data dimension reduction.**
- (d) Identify multidimensional outliers, if any, from PCA projection.**
- (e) Identify clusters and their salient characteristics from the PCA projection.**