Lab Exercises

1. Given is the case Bankruptcy given in folder Bankruptcy. The Data contains the following variables:

Name	Description
D	D=0 for bankrupt firms, D=1 for healthy firms.
YR	Year of Bankruptcy for failed firm in matched pair
R1	CASH/CURDEBT
R2	CASH/SALES
R3	CASH/ASSETS
R4	CASH/DEBTS
R5	CFF0/SALES
R6	CFFO/ASSETS
R7	CFFO/DEBTS
R8	COGS/INV
R9	CURASS/CURDEBT
R10	CURASS/SALES
R11	CURRASS/ASSETS
R12	CURDEBT/DEBTS
R13	INC/SALES
R14	INC/ASSETS
R15	INC/DEBTS
R16	UBCDEP/SALES
R17	INCDEP/ASSETS
R18	INCDEP/DEBTS
R19	SALES/REC
R20	SALES/ASSETS
R21	ASSETS/DEBTS
R22	WCFO/SALES
R23	WCFO/ASSETS
R24	WCFO/DEBTS

The variables R1,R2,…R24 are all financial ratios given. We need to find with what accuracy can we predict the Health of the firm based on this information using k – nearest neighbor algorithm. Try different values of K from 1 to 9.

2. Given is the dataset *BreastCancer.csv* in folder **Winconsin**. It is dataset of patients with breast cancer. The following are its attributes:

#	Attribute	Domain
1.	Sample code number	id number
2.	Clump Thickness	1 - 10
3.	Uniformity of Cell Size	1 - 10
4.	Uniformity of Cell Shape	1 - 10
5.	Marginal Adhesion	1 - 10
6.	Single Epithelial Cell Size	1 - 10
7.	Bare Nuclei	1 - 10
8.	Bland Chromatin	1 - 10
9.	Normal Nucleoli	1 - 10
10.	Mitoses	1 - 10
11.	Class:	benign, malignant

Fit K-NN model and find the optimal K.