MTH443: Quiz 2
Full Marks 20
Time: 17:10-18:10

Put your code, output/plot/answer/interpretation of the output in a single file and upload the pdf file through the Exam (Quiz 2) link on mooKIT.

Consider the dataset quiz2.csv having the following variables:

Variable	Description
age	Age of the patient in years
cp	Chest pain type: 0: Typical angina, 1: Atypical angina, 2: Non-anginal pain, 3: Asymptomatic
trestbps	Resting blood pressure in mm Hg
chol	Serum cholesterol in mg/dl
fbs	Fasting blood sugar level, categorized as above 120 mg/dl (1 = true, 0 = false)
restecg	Resting electrocardiographic results: 0: Normal, 1: Having ST-T wave abnormality, 2: Showing probable or definite left ventricular hypertrophy
thalach	Maximum heart rate achieved during a stress test
exang	Exercise-induced angina (1 = yes, 0 = no)
oldpeak	ST depression induced by exercise relative to rest
ca	Number of major vessels (0-4) colored by fluoroscopy
HDS	Heart disease status (0 = no disease, 1 = presence of disease)

(a) Obtain density estimate plots of the variable "chol" for HDS value 0 group and for HDS value 1 group using kernel density estimation method with a Gaussian kernel.

(b) Using the estimated density obtained in (a), find $P(chol > 250)$ for the group with HDS value 1.

(c) Split the available data into 2 parts, keeping the last 10% records as out of sample test data and apply the following classifiers to build classification models for the 2-class problem with classification variable as HDS: (i) linear discriminant function, (ii) quadratic discriminant function.

(d) Calculate the misclassification error rates of the classification models obtained in (c), separately for the learning data and the test set data.

Note: A maximum of 20 mins compensatory time to 231080042, 231080053, as per GoI rules