Coding Quiz 1

February 8, 2024

Instructions:

- Upload your .R file (Dropbox sign-in is NOT required) as per the links given below:

  BS and others:
  https://www.dropbox.com/request/I5sYRRgfBOJ2RpyCGm1D.

  MSc (231080100 - 231080108):
  https://www.dropbox.com/request/7kPoDs1EIIQGB1XPsDet.

  MSc (231080030 - 208160499):
  https://www.dropbox.com/request/XHjU15rlaecP9ccAb5hj.

  NO other files should be submitted through this link.

- Accepted format of the file: 11111111.R (where '11111111' is your roll number).

  Use your full name (e.g., Subhajit Dutta) and complete IITK email address (e.g., duttas@iitk.ac.in) while submitting the .R file.

- Output should be printed **strictly in the order of questions** given in page 2.
- When your .R file is executed, text written below in red should only be printed.
- Please **avoid spamming**, do NOT upload incorrect and/or multiple files.
- If you submit multiple files, **only the first .R file will be considered**.

  Moreover, **grading will be based solely on this .R file**.

Read the set of instructions given above again, and then go to page 2.

- Total marks: 14
- Time: 3pm to 3:30pm.

  Needless to say, .R files dropped **after 3:30pm will NOT be graded**.

- Install the R libraries **e1071, HDclassif** and their dependencies.

1. Let $a_i \in \mathbb{R}$ for all $1 \leq i \leq n$. Define the collection $\mathbb{P}_n := \{p(x) = a_0 + a_1 x + \cdots + a_n x^n : a_n \neq 0 \text{ and } x \in \mathbb{R}\}$. Consider

$$p_1 = x^2 + x + 2$$
$$p_2 = 2x^2 + 4x$$
$$p_3 = 3x^2 + 2x + 1$$
$$p_4 = 11x^2 + 13x + 4$$

be 4 polynomials in $\mathbb{P}_2$.

Find the coefficients of p_4 with respect to the basis $\{p_1, p_2, p_3\}$.

Print the coefficient vector.

2. Load the **wine** dataset.

The first column has 3 different cultivars of wine, while the remaining columns has 13 different features.

(a) Project the data on $\mathbb{R}^{13}$ to $\mathbb{R}^2$ in such a way that it captures the maximum variability present in the data.

Plot the projected data in $\mathbb{R}^2$ by assigning distinct colors to the different cultivars.

(b) Fit a *linear classifier* on the data projected to $\mathbb{R}^2$. Consider the quantity (E) defined as follows:

$$E = \frac{|y_{predicted} \neq y_{true}|}{N},$$

where $N :=$ number of observations in the response (y), $y_{true} :=$ the first column of the data and $y_{predicted} :=$ the predicted values obtained from the classifier.

Print the numerical value of E in the format "$error = E$".

Output marks: $2 + 3 + 2$

R code marks: $2 + 2 + 3$