Homework 1

The attached dataset contains the medical records of 299 patients, collected during their follow-up period, who had heart failure where each patient profile has 13 clinical features.

Thirteen (13) clinical features:

- age: age of the patient (years)

- anaemia: decrease of red blood cells or hemoglobin (boolean)

- high blood pressure: if the patient has hypertension (boolean)

- creatinine phosphokinase (CPK): level of the CPK enzyme in the blood (mcg/L)

- diabetes: if the patient has diabetes (boolean)

- ejection fraction: percentage of blood leaving the heart at each contraction (percentage)

- platelets: platelets in the blood (kiloplatelets/mL)

- sex: woman or man (binary)

- serum creatinine: level of serum creatinine in the blood (mg/dL)

- serum sodium: level of serum sodium in the blood (mEq/L)

- smoking: if the patient smokes or not (boolean)

- time: follow-up period (days)

- [target] death event: if the patient deceased during the follow-up period (boolean)

Train logistic regression model for predicting the “death event” in the following cases:

1. Using threefold split: Train/Validation/Test (70%,10%,20%)
2. Using 10-fold cross validation

In each case, report the classification accuracy