بسم الله الرحمن الرحيم

تمرین سری پنجم درس تحلیل و طراحی آماری فرایندها

Instructions:

All of previous instructions + 20% reduce score for late answers.

Explain about all of your plots.

You can use "trainControl" function for cross validation in R. caret package

You can use sample for test and train.

Use set.seed(1) in those questions are necessary.

1.

a) Use MASS::birthwt. Remove NAs. Predict bwt with age, lwt (mother's weight), smoke.(build smoke 0,1)

b) Compare model A (linear age) vs model B (ns(age, 3)) using 5-fold CV MSE. Which is better? Is there any other model you prefer?(no so sensitive and no trying every model, just any interpretation of yourself, based on any measure you use)

c) Plot partial effect of age (spline) on predicted values.(use median lwt and median smoke)

2.

a) Use 5-fold cross-validated logistic glm using V1–V9 (baseline).

b) Fit GAM: class ~ s(V1) + s(V3) + V2 + V4 + ... and compare it with the simple lm model.

c) plot gam model and interpretation.

d) Plot smooth functions for V1 and V3 (with mgcv), and 2D scatter colored by predicted prob.

3.

a) In data of pima diabetes (mlbench::PimaIndiansDiabetes), do train/test split (80/20). Fit logistic regression on suitable set: diabetes ~ glucose + mass + age.

b. Report coefficients, ORs (95% CI).

c. Predict on suitable set; confusion matrix, accuracy, sensitivity, specificity.

4. In NHANES data, Predict systolic blood pressure (BPSysAve) using BMI, Age and their interaction; include smoking as covariate. Use an 80/20 split.

B. Plot predicted BP vs BMI for two ages (30 and 60) — show interaction. (Hint: You can use for your newdata from this function of expand.grid. Much more hint: Use SmokeNow = factor("No", levels=c("No","Yes")))

5. In previous question(4), first consider visually the relationship between every covariate (just mentioned in question) with response. Based on the plot and these correlations what is your preference of the model? Fit it and interpret.

6. In aml data in Survival package:

Use 70 and 30 train test and predict time by your variables.

b) Use two gam one is simple and one with interaction and interpret them. Compare your models by suitable measures.

7. In previous question and Data, use status as your responce and fit suitable model. Interpret your OR, confusion matrix and accuracy. If any other non-linear model is your preference express it and tell why.

8. Predict birth weight using smooth terms for lwt (mother weight) and age and linear smoke. Compare GAM vs LM using 5-fold CV RMSE.

Hint: Fit GAM bwt ~ s(lwt) + s(age) + smoke and compare to lm bwt ~ lwt + age + smoke via 5-fold CV RMSE.

b. Fit final gam for plotting smoothes.

c. Use the interaction in your gam model and interpret it.