BOCCONI UNIVERSITY

Churn Prediction

20596 – MACHINE LEARNING. Final Data Challenge [Graded]
Prof. Daniele Durante

April 16, 2021

1 PROBLEM DESCRIPTION

The dataset consists of information from 5986 customers of a telecommunication company. For 4000 of these customers you know if they still have (y = 1) or not (y = 2) an active contract with the company, and you possess additional input variables describing several features of each customer and of the services subscribed. For the other 1986 customers, you have only information on the inputs, but you do not know if their contract with the company is still active or not.

Your goal is to predict y for the held–out 1986 customers.

There are **17 input variables**, which are described below.

1. `gender`: gender [Male – Female]
2. `senior`: retired customer [1 – 0]
3. `married`: married customer [Yes – No]
4. `time`: how many months the person has been a customer of the company
5. `phone`: telephone service connected [Yes – No]
6. `lines`: multiple phone lines connected [Yes – No – No phone service]
7. `internet`: customer's internet service provider [DSL – Fiber optic – No]
8. `security`: online security service connected [Yes – No – No internet service]
9. `backup`: online backup service activated [Yes – No – No internet service]
10. `protection`: equipment insurance activated [Yes – No – No internet service]
11. `support`: technical support service connected [Yes – No – No internet service]
12. `TV`: streaming TV service connected [Yes – No – No internet service]
13. `movies`: streaming cinema service activated [Yes – No – No internet service]
14. `billing`: paperless billing [Yes – No]
15. `payment`: payment method [Electronic check – Mailed check – Bank transfer – Credit card]
16. `chargesMonth`: current monthly payment
17. `chargesTotal`: total amount paid for the services for the entire time

Note that when missclassifying customers who actually will unsubscribe from the services you pay a cost of 5. Missclassifying customers who will not unsubscribe is less dangerous and costs 1.