

リーダーボードを見ることができなくなり順位不明の JustVisiting です。自分が知る限りなんとも中途半端な結果だったと思うのですが（ぎりぎり）上位入賞候補だということでご連絡をいただき、なんともありがたい限りです。そんなわけで、これを読むのはほぼ TA の方くらいだろうと思って書いている次第です。TA の方々、お疲れさまです。コードがぐちゃぐちゃですみません。もしこれを読んでいる勤勉な受講生がいるとしたら、お読みいただきありがとうございます。まだ終わっていなければ、最終課題頑張ってください。分量だけ多く内容も面白みもない文章になってしまいましたが、少しでも参考になれば幸いです

今回のコンペは、やや締め切りが近づいてから取り組み始めたのもあって、全体としてやや力任せな感じになってしまいました。まずデータについて、初心者なもので相対的なことはわかりませんが（タイタニックもほぼ beginner を提出しただけで終わったので）、あまり質が良くない印象でした。Np.nan と XNA が混在する中、Ext_source など、複数のカラムで多くの欠損値があり、また数値データの分布がほぼ全て不自然なものになっていました。加えて own_car_age、days_employed に明らかに異常な外れ値がかなりの割合あったり、day_last_phone_change に 0 が大量にあったり、Male の widow がいたり、、、と、これ以上はただの愚痴になりそうなのでやめておきますが思わず突っ込みたくなるような記録が大量にありました。これらを丁寧にきれいにするのが良かったのだと思いますが、なにせ時間がなかったもので、わかっていながらあまり手を加えられなかったのが残念です。また work の mobile だとか home の phone だとか、OBS とか DEF とかの DPD だとか、意味のよくわからないカラムが多く苦労しました。はじめはとりあえず欠損値をそのまま扱う lightgbt, xgboost にそのまま丸投げしましたが、それでも 0.74 くらいのスコアが出ました。金額系のカラムが重要そうで、欠損が少なく、かつ理解可能だったので lightgbt で欠損を埋め、さらに引いたり割ったりして特徴量を増やしました。特に credit/annuity から、おそらく返済期間であるものが算出できたのは効いたと思います。他にも数値系は軒並み意味がわかる限り割ったり引いたりして特徴量を増やしました。文字のエンコーディングは安直に意味のない整数にしましたが学歴だけは上下関係をつけました。この時点で変数増加法により特徴量を選択して lightgbm で 0.76 弱、xgboost で 0.75 くらい、特徴量すべて突っ込んで同 0.75、0.74 くらいでしたが、結局 optuna でチューニングしたところ、すべて 0.76 弱にはなりました。余談ですが（今回は）xgboost はそのまま扱うと lightgbt に比べて若干スコアが劣り実行はかなり遅かったのですが（自分は colab で作業しました）GPU を使うと lightgbt よりも早くなりましたし双方チューニングするとスコアは逆転しました。また特徴量が多いほどチューニングの効果が出る印象でした。ですが、毎回ノッてきたところで google にこれ以上 GPU を使うなら金を払えと言われて止められるもので、はじめの感動とは対象的に正当性のかけらもない不満が溜まったこと事実です。一応、欠損値を除いて他のモデルも試したところ、svm は何時間たっても終わらず、ロジスティック回帰、knn、mlp、randomforest は 0.6 ~ 0.7 くらいでした。ここから、これが正しいやり方かどうかはわかりませんがこれまで作ったモデルをすべて 5 分割ですたきんぐ？して、それぞれの予測値でできた新たなデータセットを作成しました。このデータセットでの学習では線形モデルのスコアが明らかに良かったので、mlp、ロジスティック回帰、xgboost の gblineair を使い、0.766 くらいのスコアになりました。もう一段くらい増やす予定でしたが mlp において、それぞれのフォールドではいいスコアが出るものの、全体を結合するとスコアが大きく下がるのがわかり、困り果てているうちに時間切れとなりました。原因はよくわかりません、どなたか教えて下さい。ここまで書いておきながら、omnicampus でスコアを確認したところ 0.76115 になっていて、いまさら今回の連絡がなにかの間違いだったのではないかと不安になってきたので、書くのはここまでにしておきます。ダラダラとし

た文章をここまで辛抱強く読んでくれた TA の方、(もしいるのならば受講者の方) ありがとうございます。これからもよろしくお願いします。