

1. ソースコード

学習部分：weight.py

予測部分：simplemodel.py

simplemodel.pyをそのまま実行すれば再現できる

2. モデリング詳細

- 移動平均と似ている考えで、下記の式で予測値を計算した

$$f = w_1 * x_1 + w_2 * x_2 + w_3 * x_{12} + w_4 * 1/9(x_3 + \dots + x_{11})$$

- 重みは過去のデータから学習
- 欠損値のある予測部分に対しては別対応した。
- 後は細かい調整も行った

3. 外部データ（該当者のみ）

分析の段階で、「1世帯当たり月別のお菓子支出金額」（総務省家計調査報告）を利用した。（添付ファイル：h2803.pdf）

4. 表彰式の参加可否

まだ分からない

5. お住いの都道府県

東京

6. アンケートにご協力下さい

DeepAnalyticsはどこでお知りになりましたか？

去年1回目の時は先生からのお知らせメール

今回のコンペにおける実質の作業時間はどのくらいでしたか？

15時間ぐらい

今回のコンペについてご感想／ご意見ございましたらお願い致します。

訓練データが少ないことに対して、自分なりに苦労したが、今回はうまくいかなかったようだ。過学習のことが予想されたけど、そんなにひどい過学習とは思わなかった（笑）。

今回のコンペを通じて、自分の知識不足という点に痛感した。予測を行う時、データとちゃんと向き合って、直感でアイデアを出すのがもちろん重要だけど、業務知識および数理知識も不可欠である。

業務知識について、どういう要素が販売数に関連しているか（競合他社の動き？食生活の変化？観光客？気候？）、どんなデータがそれらの要素を反映できるか（SNS？家計調査？決算報告書？）。今回はいい答えが見つからなかった。

数理知識について、売上（販売数）を予測の時、いくつかの定番がある（多項式回帰、ARMA、マルコフ、移動平均など）が、どちらも詳しくない。また、訓練データの数が極端に少ない時どう対応すればいいのか、大きい外れを出たくない時どうすればいいのか。このようなデータサイエンティストとして基本的な知識に対して、知らないのが恥ずかしかった。

今回は前回と少し違い、実際のビジネス問題を触れ、データ分析を実際のビジネスに行かせる経験をいただいて、本当に勉強になった。特に予定がなければ、来年も参加する予定で、その時はまたよろしくお願いします。