Search kaggle

Competitions Datasets Kernels Discussion Jobs





Walmart Recruiting II: Sales in Stormy Weather

Predict how sales of weather-sensitive products are affected by snow and rain

485 teams · 2 years ago

Overview

Data

Discussion

Q

Leaderboard

More

My Submissions

New Topic



threecourse

1st place

First Place Entry

posted in Walmart Recruiting II: Sales in Stormy Weather 2 years ago





Thank you all people around this competition, I'm a newbie in data science and it was the first challenge for a non-playground competition, so I'm really surprised and glad to win.

I'm not great at English, so wrote this method description in itemized style.

Train model

- 1. Exclude item/stores whose units are all zeros.
- 2. For each item/stores, apply curve fitting by R ppr function (projection pursuit regression). $y = log1p_units, x = days from 2012-01-01$

here, data on 2013-12-25 are excluded. (because units are almost all zeros)

3. Train linear model with lasso using vowpal wabbit.

y = log1p_units - ppr_fitted

features:

- A: weekday, is weekend, is holiday, is holiday and weekday, is_holiday_and_weekend

- B: item nbr

- C: store nbr

- D : date

- E: year, month, day

- F: is_BlackFriday-3days, -2days, -1day, is_BlackFriday, +1day, +2days, +3days

- G: weather features (is preciptotal > 0.2, depart > 8, depart < -8)
- interactions A*B A*C B*E C*E B*F C*F

here, below are excluded:

- on 2013-12-25
- moving average(21 elements, centered) is zero.
- 4. Mark dates as "too much zeros" where both sides are many successive zeros.
- 4-1. for dates whose units are not zero, calculate minimum of both side successive zeros (= min_side_zeros).
- 4-2. for each item/stores,

calculate maximum of min_side_zeros (= max_min_side_zeros), floor and ceiling by 1 and 9.

4-3. for each item/stores,

mark dates as "too much zeros" where both sides are successive zeros more than max_min_side_zeros.

Prediction on test set

predicted_log1p = ppr_fitted(train-2) + linear model predicted(train-3)
predicted = exp(predicted_log1p) - 1

here, below are predicted as zero.

- item/stores whose units are all zeros.
- on 2013-12-25
- moving average(21 elements, centered) is zero.
- "too much zeros" (train-4)

Comments

The core idea is very simple like that:

- 1. Create a baseline for each item/stores.
- 2. Apply linear regression using vowpal wabbit with many features.

As for baseline:

- R ppr functions fit really nice on almost all item/stores (can be improved on some item/stores).
- At first I used moving average. It worked, but fluctulates too much or catch too distant value.

As for features:

- weekday is the most important
- month periodicity is on some store/items
- around Black Friday sales fluctuates a lot
- weather features are not effective almost at all
 In the data, people go shopping as usual however much it rains.
 It's not natural, so I guess weather data came from different stations.

Considering successive zeros was my final push, it slightly improved the score.

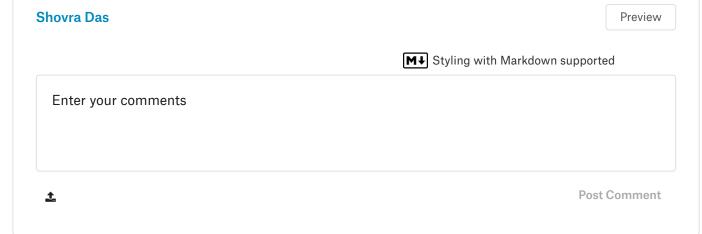
Codes

uploaded on github, https://github.com/threecourse/kaggle-walmart-recruiting-sales-in-stormy-weather

Options

Comments (5) Sort by Hotness







threecourse • (1st in this Competition) • 2 years ago • Options • Reply





Replaced with a small value or zero.

Weather data has little effect, so I believe any method will be the same.



Dmitry Larko • (2nd in this Competition) • 2 years ago • Options • Reply



Great approach! Thank you for sharing!



T. Scharf • (3rd in this Competition) • 2 years ago • Options • Reply



nice work - thanks for sharing



Vijay • (289th in this Competition) • 2 years ago • Options • Reply



@threecourse: How did you deal with missing data from weather file?



Yong Jiang • 6 months ago • Options • Reply



Thanks for sharing.

This is very impressive. When I first time looked at the weather data and found it has no correlation to the items sold. So I thought it is more dependent on week days or holiday. Your approach proves I was right. :)

© 2017 Kaggle Inc

Our Team Careers Terms Privacy Contact/Support