Walmart Stormy Weather Competition

1st threecourse

* Baseline model using Projection Pursuit Regression (non-linear smoother)
* Model deviations from baseline using L1-regularized linear model with Vowpal Wabbit
* Features:
  + Weekend/weekday, holiday and interactions
  + Item, Store
  + Date, Year, Month, Day
  + Blackfriday indicator (with +-3 day window)
  + Weather indicators (preciptotal > 0.2, depart > 8, depart < -8)
  + Interaction effects

3rd T. Scharf

* + Features
    - Weather
      * 7 MA of relevant weather (Forward and backward)
    - Date (Month, DoW, DoY, DoM, Trend) (encoded numerically)
  + Ensemble of 1000 XGB’s with random HP’s
  + Stacking using XGB
    - Level 1 preds + Day + Month
* 5th Korzekwa
  + Local Gaussian Process Regression
  + Weighted Avg of two GP’s
  + Features
    - Trend, Dow, Dom, Quarter, Start of Month
    - Storm tomorrow and no storm today and previous 3 days
* 6th Little Boat
  + Leaderboard for feedback
  + Ensembling (Blending using Ridge)
    - Linear regression (Ridge, OLS)
    - RF (global)
    - SVM
    - TS Models (ARIMA) – around 0.11 on lb
  + 2 feature sets
    - Date info + weather
    - Date info + historical sales (Last Year – day, week, month avg)
* 11th vtKMH
  + Features
    - Rolling-features (30, 14, 7, 5, 3)
      * Mean, Median, Max, Sd, Pct zeros, mean of non-zero, median of non-zero
      * Forward and backward
    - Month, DoW,
    - Weather
      * Including Lead effects
  + Ensemble
    - 6 Boosted Trees
    - 4 ExTrees
    - 1 KNN
    - 1 LM
  + Few manual adjustments
* 19th Turkewitz
  + LASSO
    - Dow
    - MA 7 (backward and forward)
    - Start/end of month

Time series models?

* 6th reports around 0.11 on lb using TS (ARIMA)