Report for CMU Data Science Cup 2016

Team The Third Place

I. Findings

II. Methdology

III. Baseline

We have two baseline models used for comparision. The first one is just pure chance, randomly select a half to be true and a half to be false as the response vector.

The second baseline is trying the probability that each household purchasing eggs as a bernoulli and we have a binomial distrbution in the case and the conditionay probabilit is given by:

$$P\{GET\ EGGS\} = \frac{\sum_i^{household} \sum_j^{day} GET\ EGGS_{ij}}{\sum_i^{household} \sum_j^{day} 1}$$

$$\begin{aligned} P\{GET\ EGGS|\ week\} &= 1 - P\{GETNOEGG\} \\ &= 1 - (1 - P\{GET\ EGGS\})^7 \\ &= 0.126284 \end{aligned}$$