

Restaurant Rating Adjustment on Yelp from User Review

Kenny Sung

Yin Lin

Chih-Ang Wang

ECE 18-799: Evolutionary Algorithms in
Engineering Optimization
Fall 2014 Course Project

Project Description



Project Description

UNDERSTANDING ONLINE STAR RATINGS:



Project Goals

- Adjust restaurants' star ratings by using restaurants' attributes and users' review

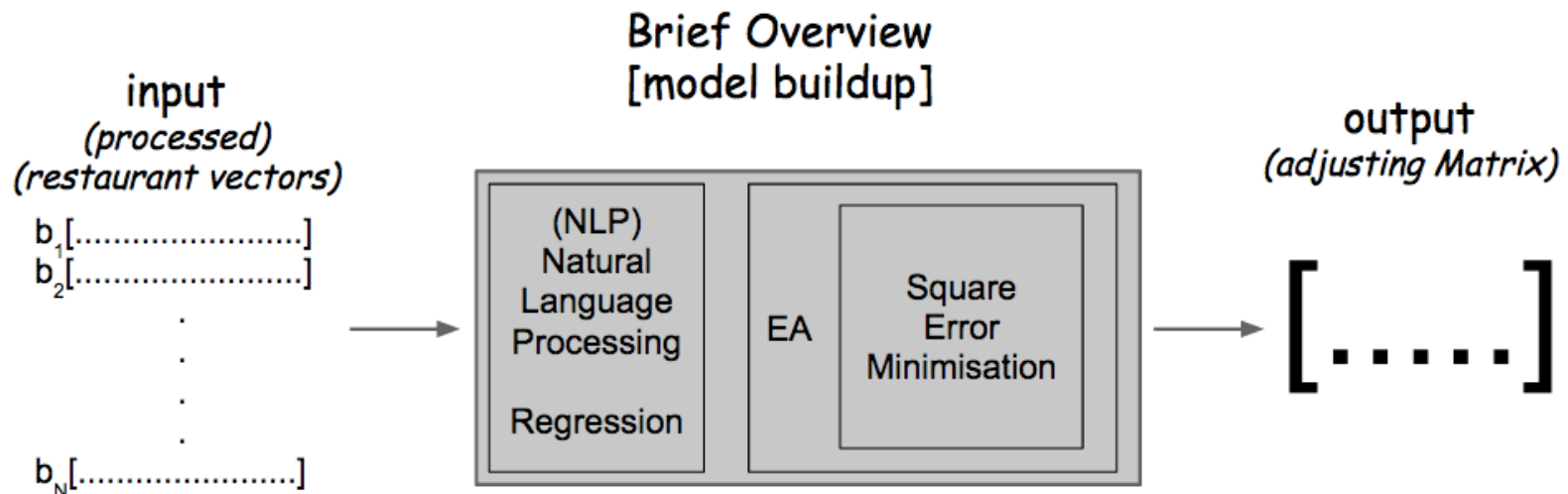


Background

- Related Work
 - User review visualization
 - Restaurant improvement from user reviews
- Related Techniques
 - Natural Language Processing
 - Feature Selection
 - Evolutionary Algorithm
 - Linear Regression

Application

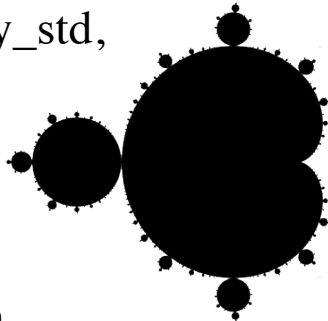
A star rating model based on Yelp dataset, which provides a more accurate star rating to reflect the true quality of a restaurant



Application

- [restaurant vector]

| | |
|---------------------------|------------|
| [review_polarity_mean, | [0.32202, |
| review_polarity_std, | 0.22073, |
| review_subjectivity_mean, | 0.58396, |
| review_subjectivity_std, | 0.12571, |
| vote_cnt_mean, | 2.96313, |
| vote_cnt_std, | 5.78322, |
| likes_mean, | 0.0, |
| likes_std, | 0.0, |
| tips_polarity_mean, | 0.29853, |
| tips_polarity_std, | 0.40612, |
| tips_subjectivity_mean, | 0.39373, |
| tips_subjectivity_std] | 0.36093] |



TextBlob

Fitness Function

$$f(\mathbf{w}) = Error(\mathbf{w}) = \sum_{\mathbf{x} \in X} \left\| \boxed{y} - \boxed{\mathbf{w}^T \mathbf{x}} \right\|^2$$

$\mathbf{x}, \mathbf{w} \in \mathbb{R}^d$

Ground Truth Ratings

Estimated Ratings

\mathbf{w} : model (regression coefficient); individual in ES
 \mathbf{x} : training example, i.e. feature vector of a restaurant
 y : ground truth rating for a restaurant

Genetic Algorithm

| | |
|-----------------------|-------------------------------------|
| Representation | Real-valued vectors |
| Mutation | Uniform mutation |
| Recombination | Single arithmetic recombination |
| Parent Selection | Stochastic universal sampling (SUS) |
| Survivor Selection | Elitism |
| Termination condition | 300 generations |
| Cross validation | 3-folds |

Experimental Setup

[Representation] (40 individuals)

$\langle x_1, x_2, x_3, \dots, x_{12} \rangle$

↖ object variable

init(-1 ~ +1)

[Mutation] ($p_m = 1/12$)

$x_i' = x_i + N(0, 5)$

[Crossover] ($p_c = 0.7$)

$\langle x_1, \dots, x_{k-1}, \alpha y_k + (1-\alpha)x_k, x_{k+1}, \dots, x_n \rangle$

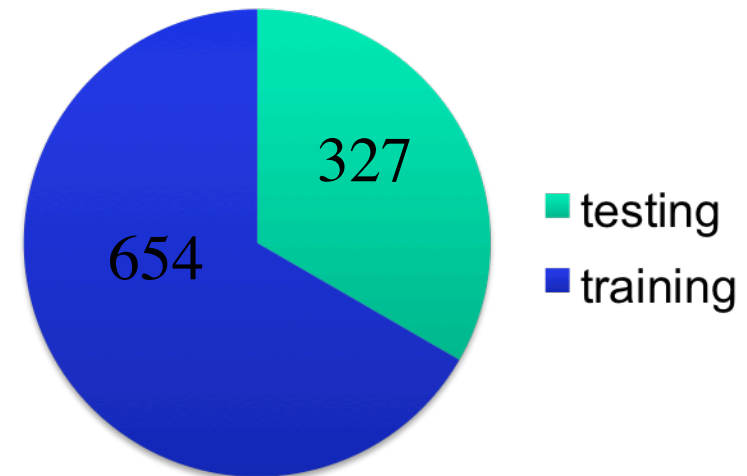
$\langle y_1, \dots, y_{k-1}, \alpha x_k + (1-\alpha)y_k, y_{k+1}, \dots, y_n \rangle$

$\alpha = 0.5$

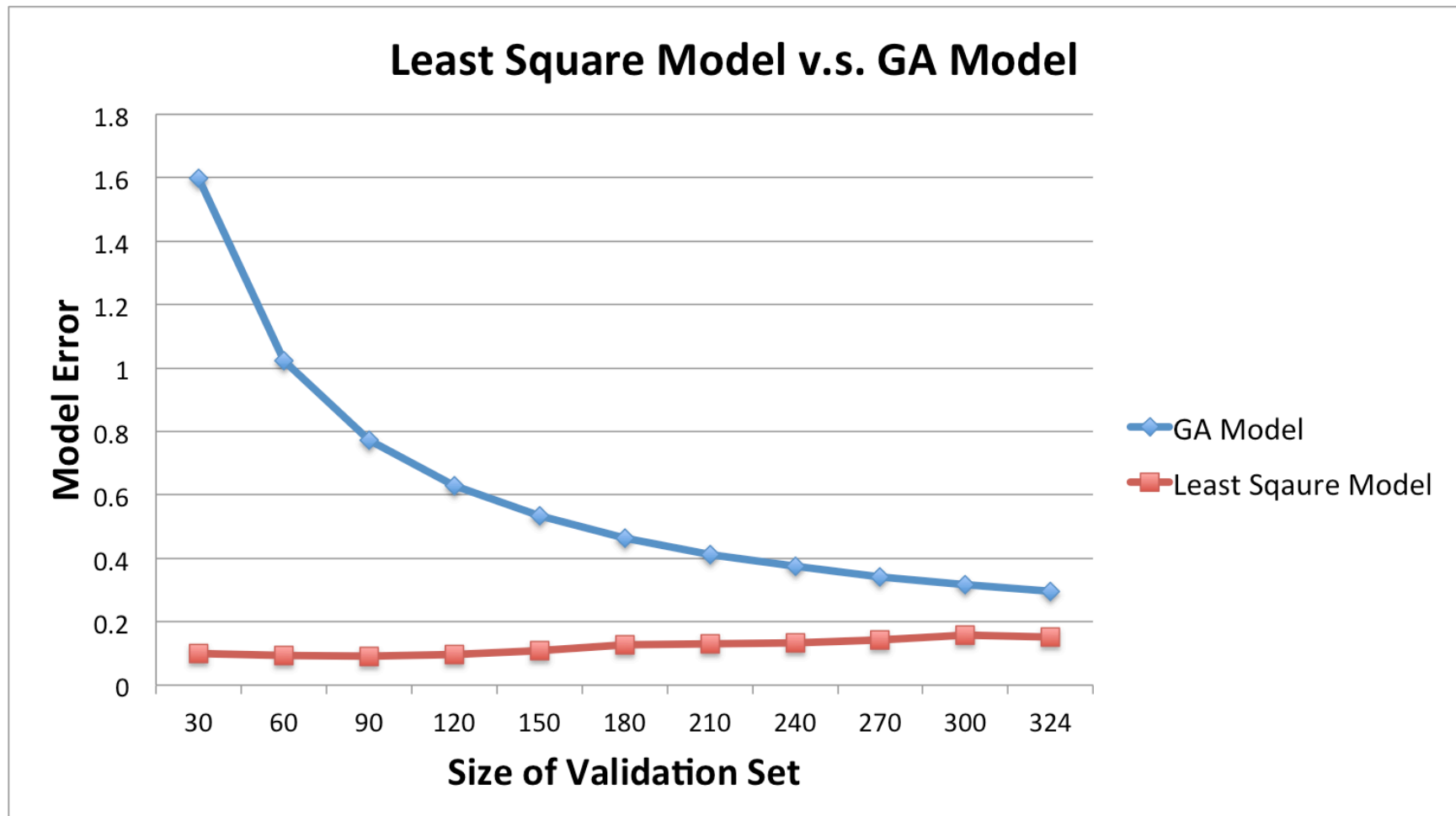
[Survivor Selection]

50% with elitism

[Cross Validation] (3-folds)



Results for GA



Evolutionary Programming

| | |
|-----------------------|---|
| Representation | Real-valued vectors |
| Mutation | Gaussian Perturbation with self-adaption |
| Recombination | None |
| Parent Selection | Everyone breeds ! |
| Survivor Selection | $(\mu + \mu) = (40 + 40, 40)$ by tournament |
| Termination condition | 300 generations |
| Cross validation | 3-folds |

Experimental Setup

[Representation] (40 individuals)

$\langle x_1, x_2, x_3, \dots, x_{12}, \sigma_1, \sigma_2, \sigma_3, \dots, \sigma_{12} \rangle$

↖ object variable

init(-5 ~ +5)

↖ strategy parameter

init(1)

[Mutation] ($\alpha=0.2, \varepsilon=0.1$)

$\sigma_i' = \sigma_i (1 + \alpha * N(0, 1))$

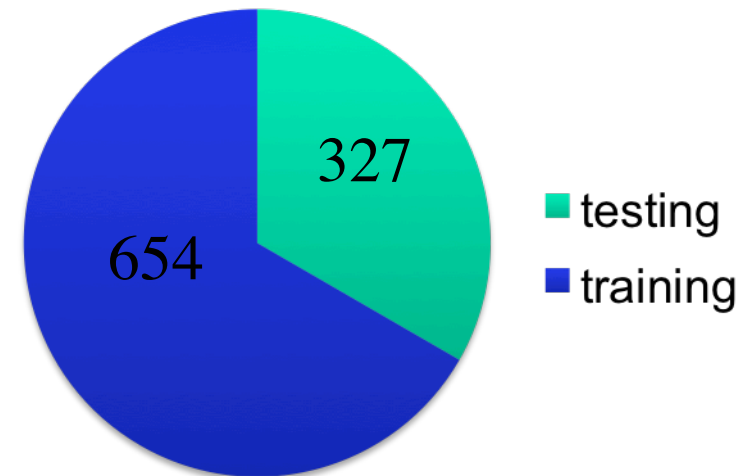
$\sigma < \varepsilon \Rightarrow \sigma = \varepsilon$

$x_i' = x_i + \sigma_i' N_i(0, 1)$

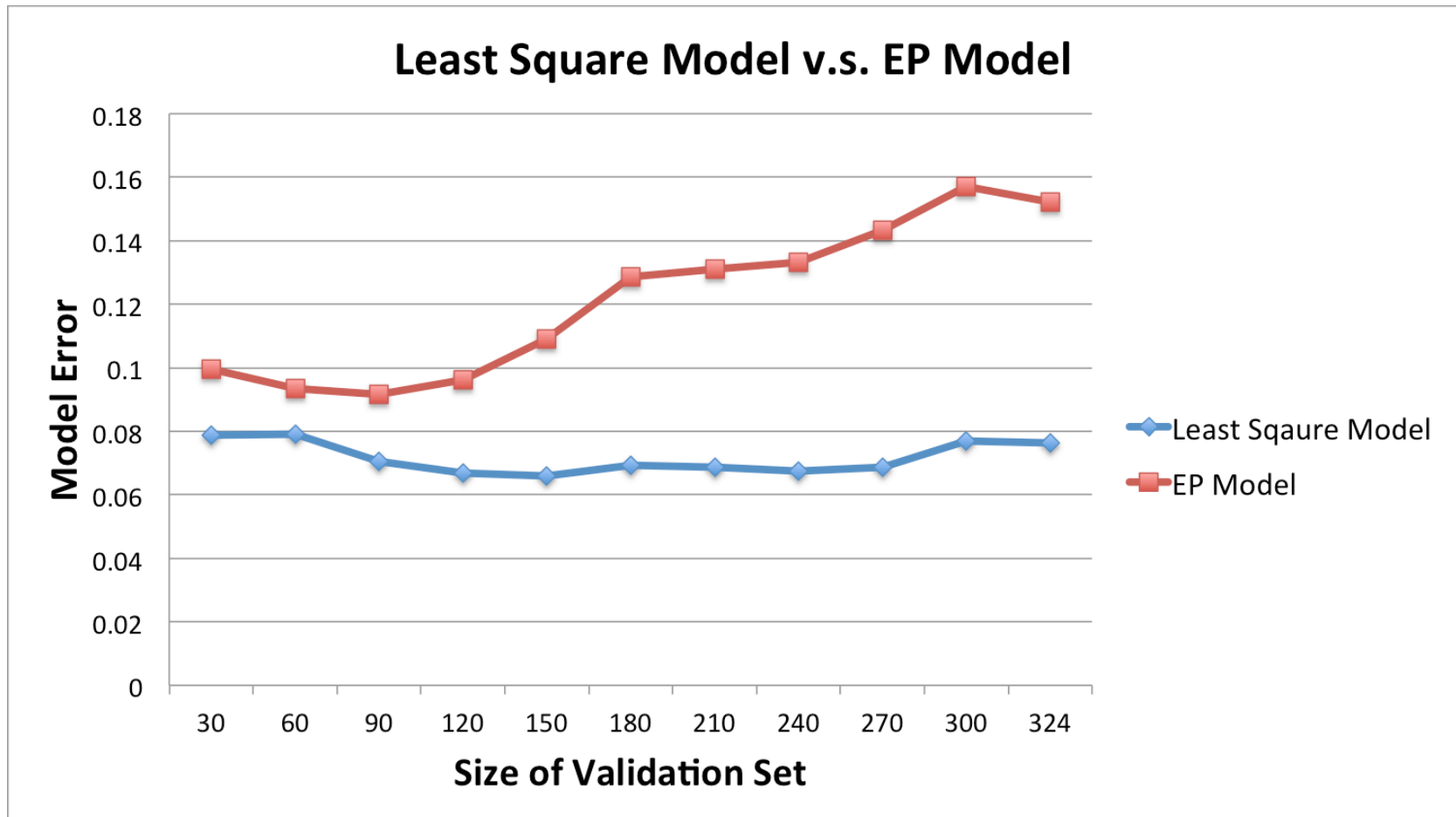
[Survivor Selection] ($q=10$)

50% with tournament selection

[Cross Validation] (3-folds)



Results for EP



Evolutionary Strategies

| | |
|-----------------------|---|
| Representation | Real-valued vectors |
| Mutation | Gaussian Perturbation with self-adaption |
| Recombination | Object variables: Global Discrete Mutation parameters: Global Intermediate |
| Parent Selection | Uniform Random |
| Survivor Selection | $(\mu + \lambda, \mu) = (40+300, 40)$ |
| Termination condition | 300 generations |
| Cross validation | 3-folds |

Experimental Setup

[Representation] (40 individuals)

$\langle x_1, x_2, x_3, \dots, x_{12}, \sigma_1, \sigma_2, \sigma_3, \dots, \sigma_{12} \rangle$

↖ object variable

init(-5 ~ +5)

↖ strategy parameter

init(1)

[Mutation] ($\tau = 0.02, \tau' = 0.038$)

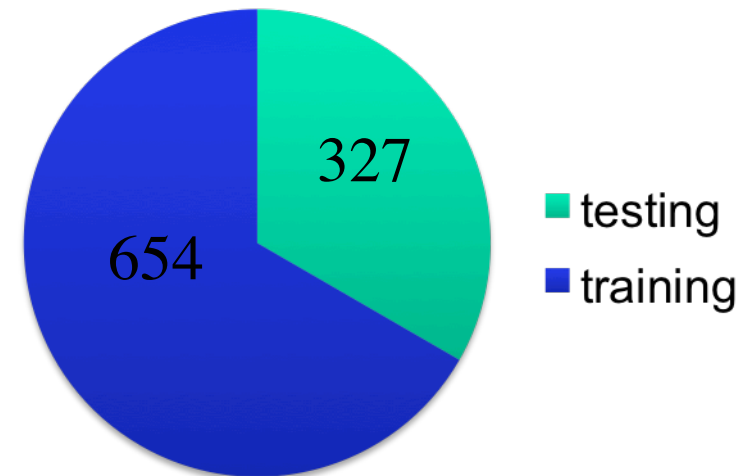
$\sigma_i' = \sigma_i \exp(\tau N(0, 1) + \tau' N_i(0, 1))$

$x_i' = x_i + \sigma_i' N_i(0, 1)$

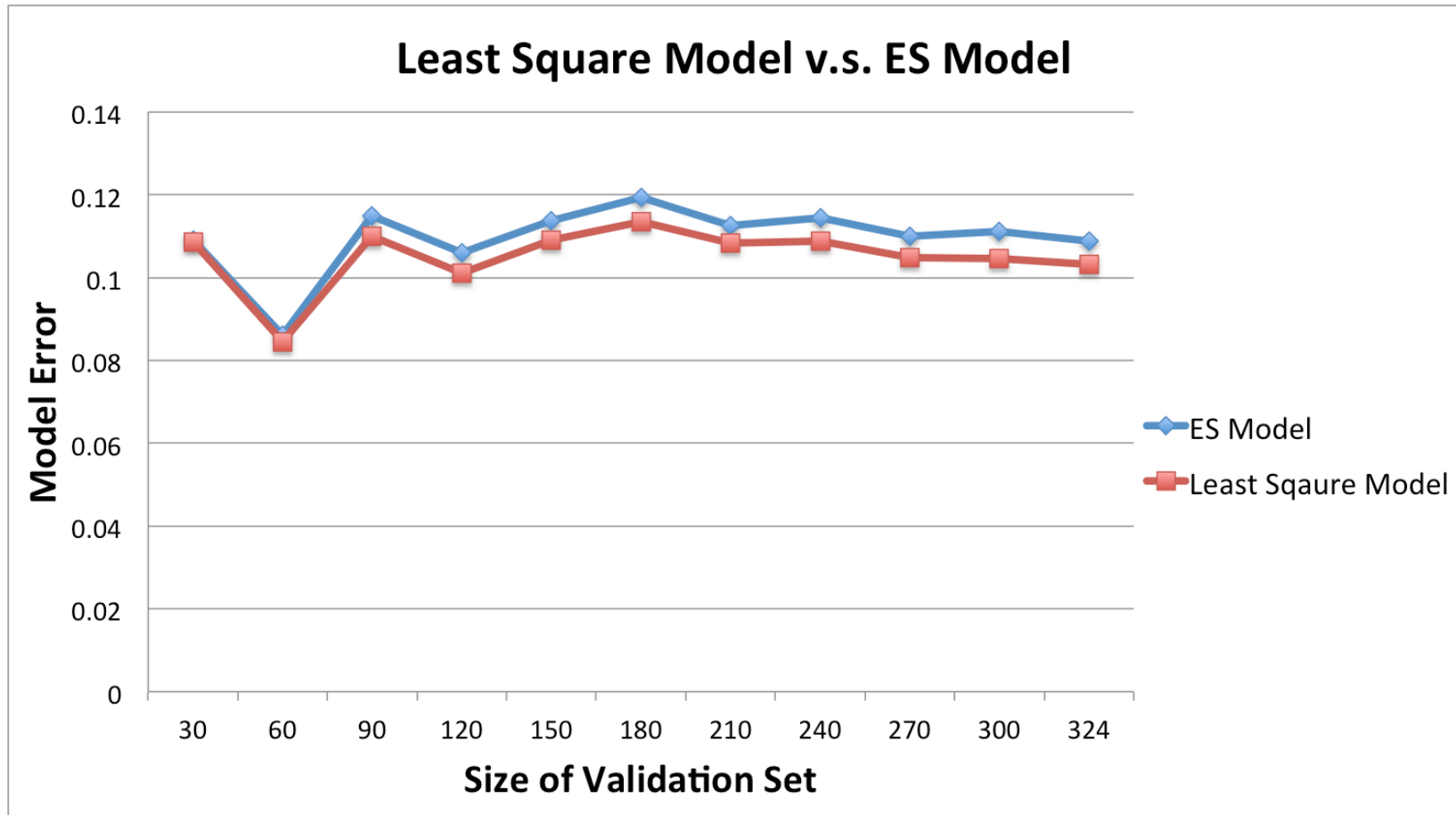
[Survivor Selection]

$(\mu + \lambda, \mu) = (40 + 300, 40)$

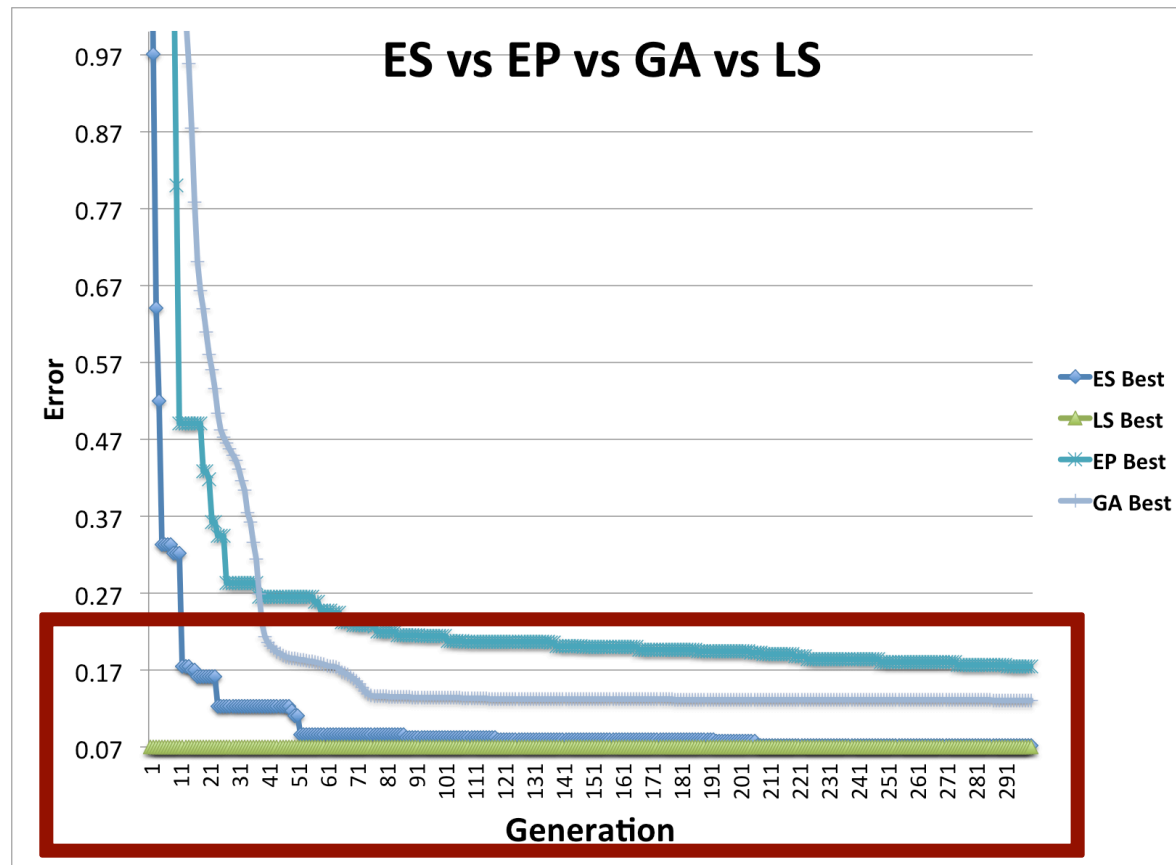
[Cross Validation] (3-folds)



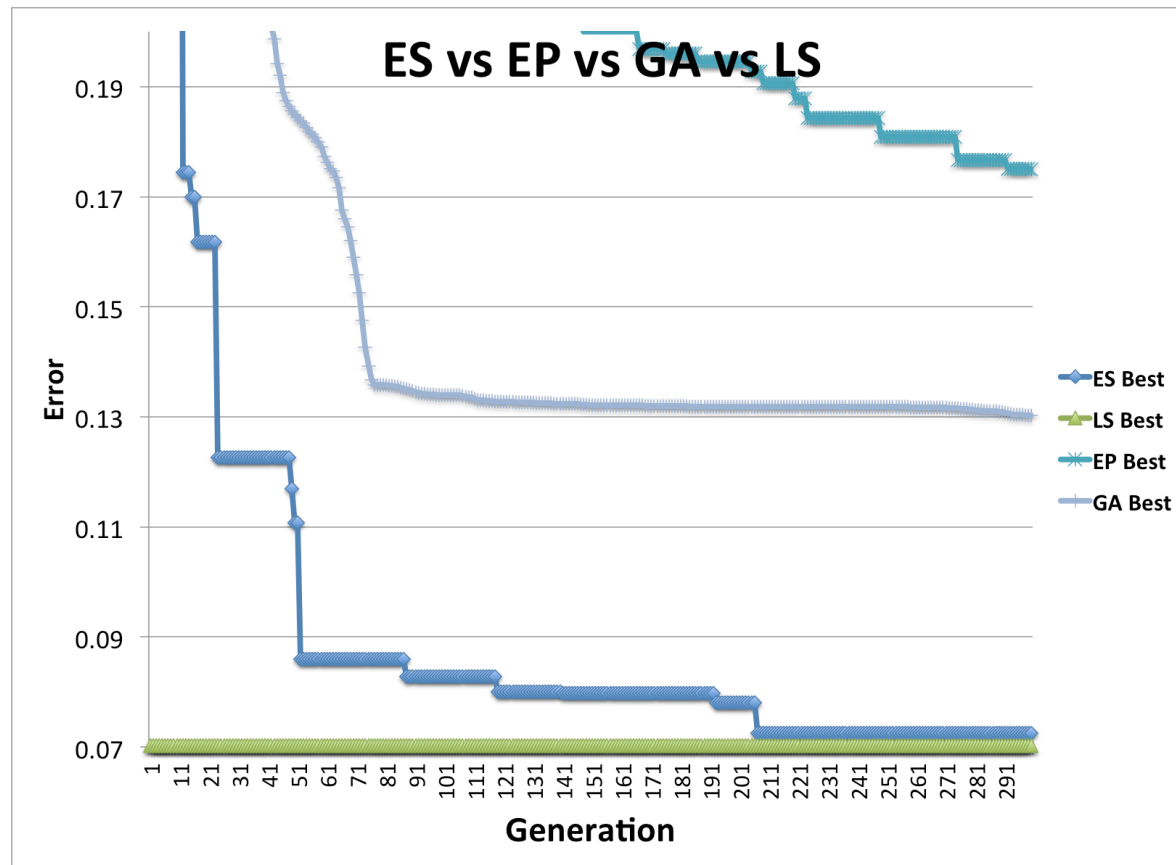
Results for ES



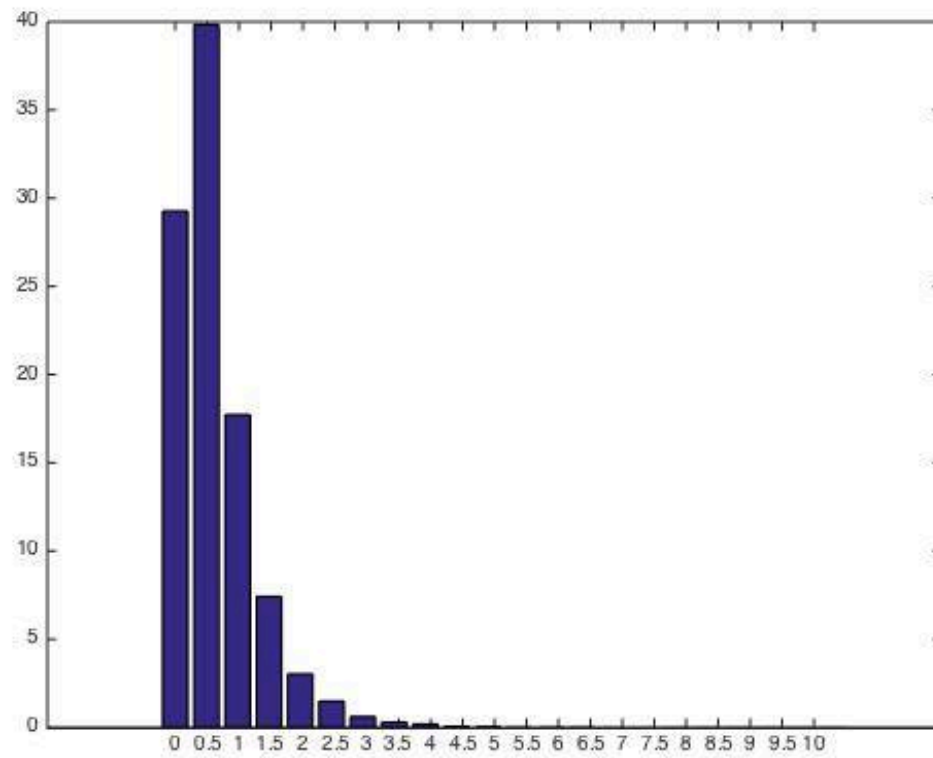
Overall Results



Overall Results



Rating Re-Adjustment



Conclusion and Future Work

- Conclusion
 - Converge Speed: $ES > GA > EP$
 - AVG(Square Error): $ES < GA < EP$
 - 70% restaurants on Yelp need to be adjusted!
- Future Work
 - Neural network, non-linear model may provide better performance!