Prediction of Cellular Service Quality

MAXMILLIAN MINICHETTI[1]

maxmin@stanford.edu | [1] Department of Electrical Engineering | Stanford University

Introduction

Objective: To predict the Received Signal Strength Indicator (RSSI) across mobile devices for gauging Quality of Experience (QoE)

Background:

- RSSI generally negative with 0 corresponding to the strongest possible signal (measured in dB)
- Typically, supervised learning uses RSSI to obtain meaningful geolocation insights for optimizing wireless sensor networks (direction-finding, indoor sensor localization, noise reduction)
- Predicting RSSI from noisy location readings is a topic far less explored (adding basic device info and network data can help)

Ordinary Least Squares Regression

Method:

$$J(\theta) = \|X\theta - y\|^2 = \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

$$h_\theta(x) = \sum_{i=0}^{n} \theta_i x_i = \theta^\top x \qquad \theta^* = \left(X^\top X\right)^{-1} X^\top y$$

Baseline Model:

- Includes all first-order covariates pulled directly from training set
- Provides intuition for judging the significance of each regressor

Model Selection Strategies:

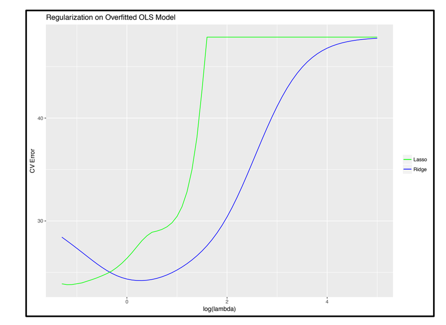
- Added higher-order terms and interactions
- **Forward Search** using the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) to penalize model complexity

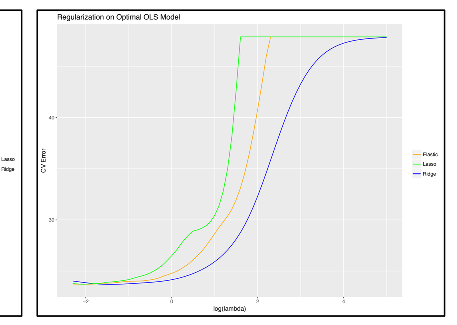
$$AIC = \frac{n}{\hat{\sigma}^2}\left(\mathrm{Err}_{\mathrm{tr}} + \frac{2|S|}{n}\hat{\sigma}^2\right)$$

$$BIC = \frac{n}{\hat{\sigma}^2}\left(\mathrm{Err}_{\mathrm{tr}} + \frac{2|S|\ln(n)}{n}\hat{\sigma}^2\right)$$

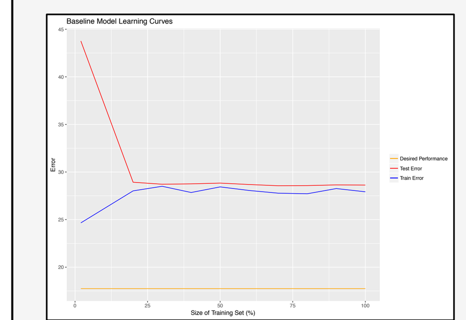
- **Elastic Net Regularization** to further combat overfitting and identify the most explanatory covariates

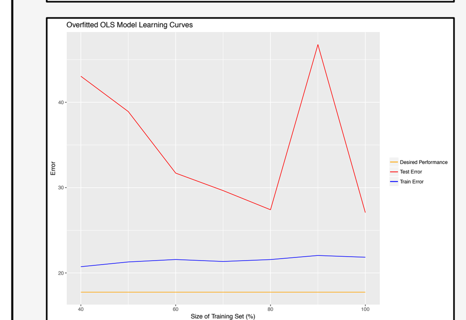
$$\begin{aligned} J_{EN}(\theta) &= \|X\theta - y\|^2 + \lambda\left(\alpha\|\theta\|^2 + (1-\alpha)\|\theta\|_1\right) \\ &= \sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2 + \lambda\left(\alpha\sum_{j=1}^{n}|\theta_j|^2 + (1-\alpha)\sum_{j=1}^{n}|\theta_j|\right) \end{aligned}$$



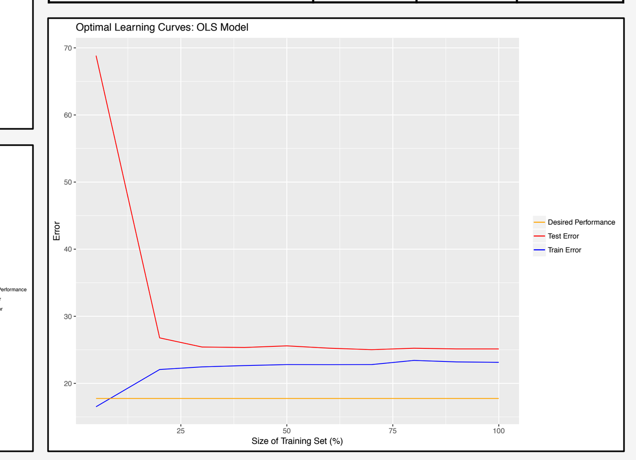


Results:





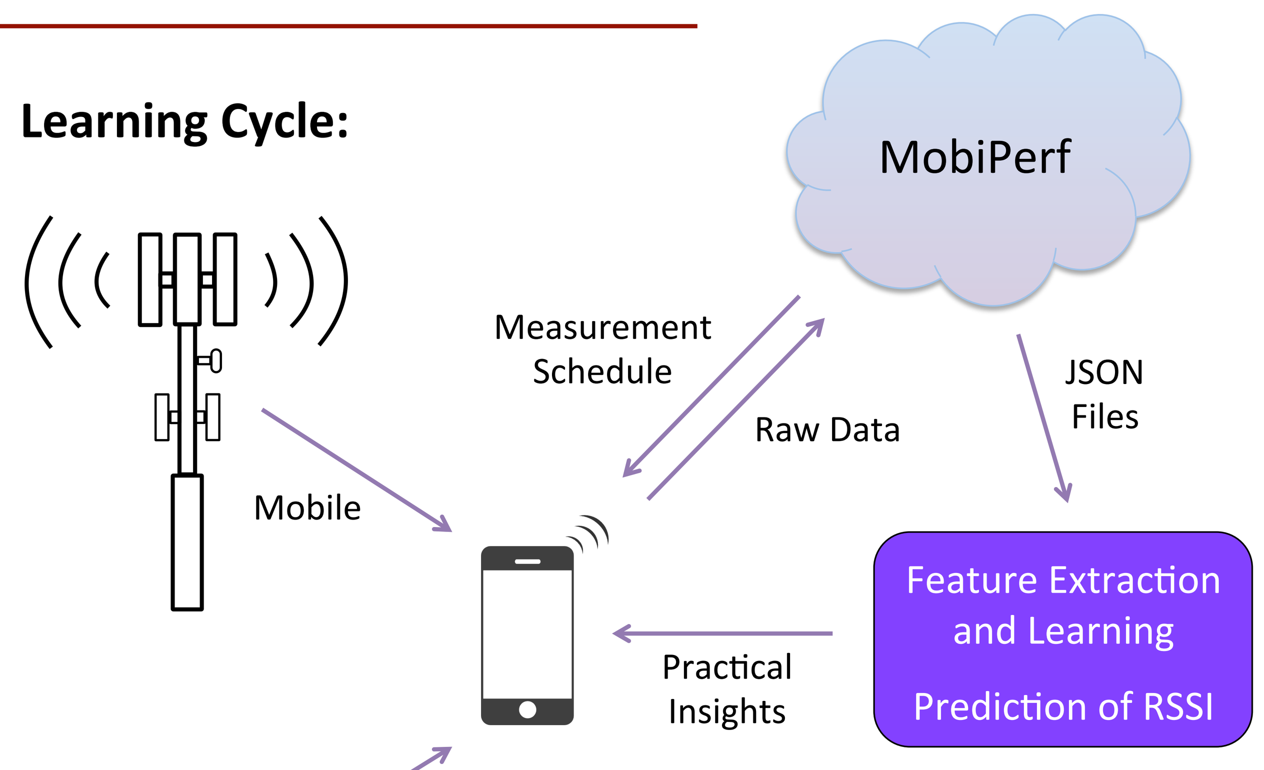
Model \ Error Metric	Train	CV	Test
Baseline	27.921	28.040	28.625
All Covariates	21.842	44.570	27.076
Forward Search (AIC)*	23.135	24.157	25.129
Forward Search (BIC)	24.270	24.695	26.154



Dataset and Features

- MobiPerf is an open source application that monitors network performance across mobile devices
- Network measurements from a specific device are collected throughout the day
- Each log tracks latency, bytes transferred, and other relevant attributes of mobile devices sending queries through ping, traceroute, and HTTP requests
- Focus on ping requests, as they are most relevant to network connectivity testing
- Used logs from Nov. 1 to Nov. 15, 2016
- Compiled data into single data frame of nearly 9,000 samples
- Set aside 20% for model testing

Learning Cycle:



Relevant Features:

Battery Level, Charging Status, Network Type, Cell Carrier, Device Model, Manufacturer, Latitude, Longitude, Data Packets Sent, Packets Lost, and Time of Day

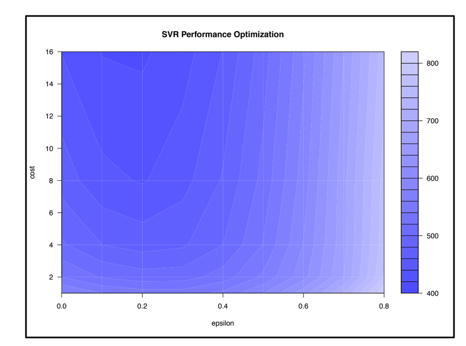
Support Vector Regression

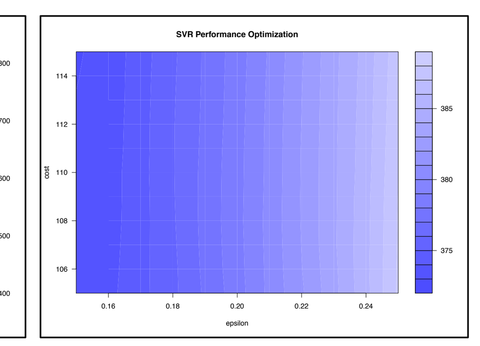
Method:

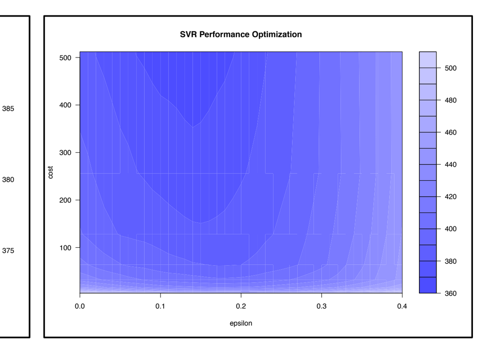
$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|\theta\|^2 + C\sum_{i=1}^{m}(\delta_i^- + \delta_i^+) \\ \text{subject to} & \begin{array}{lcl} y^{(i)} - \theta^\top x^{(i)} - b & \leq & \epsilon + \delta_i^- \\ \theta^\top x^{(i)} + b - y^{(i)} & \leq & \epsilon + \delta_i^+ \\ \delta_i^-, \delta_i^+ & \geq & 0 \end{array} \end{array}$$

$$\Downarrow$$

$$f(x) = \sum_{i=1}^{m}(\lambda_i - \lambda_i^*)\, K\left(x^{(i)}, x\right) + b$$





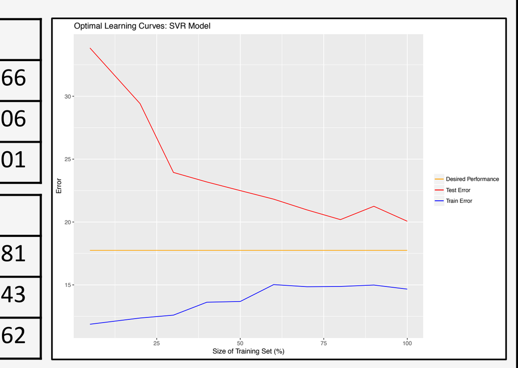


SVR Tuning: Many SVR models were trained over the parameter space to obtain optimal values for cost (C) and epsilon (ε) via 10-fold cross-validation

Results:

Kernel Function	Train	CV	Test
Linear (C = 1, ε = 0.1)	29.194	29.223	29.766
Poly (C = 1, ε = 0.1)	23.819	32.784	27.606
RBF (C = 1, ε = 0.1)*	22.537	23.677	24.601

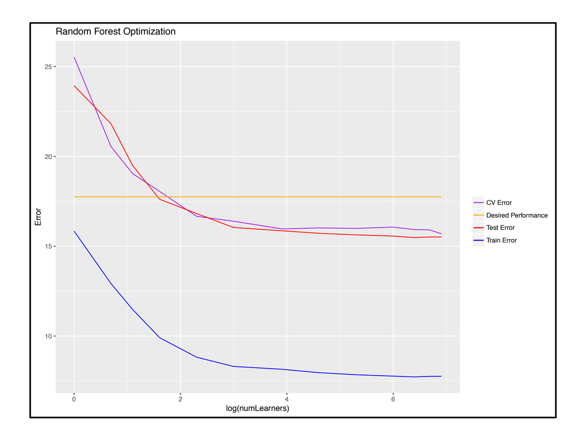
Parameters	Train	CV	Test
RBF (C = 110, ε = 0.2)	16.493	19.691	20.581
RBF (C = 115, ε = 0.15)	16.320	19.617	20.643
RBF (C = 512, ε = 0.14)*	14.665	19.182	20.062



Random Forest Regression

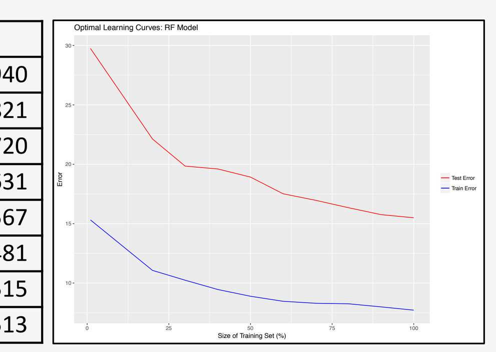
Method:

- RF algorithm fits a number of randomly generated decision trees (of a specified maximum depth) on bootstrapped subsets of the original data
- Improved model by tuning the number of learners
- Tree bagging reduces variance by averaging over many complex models with low bias



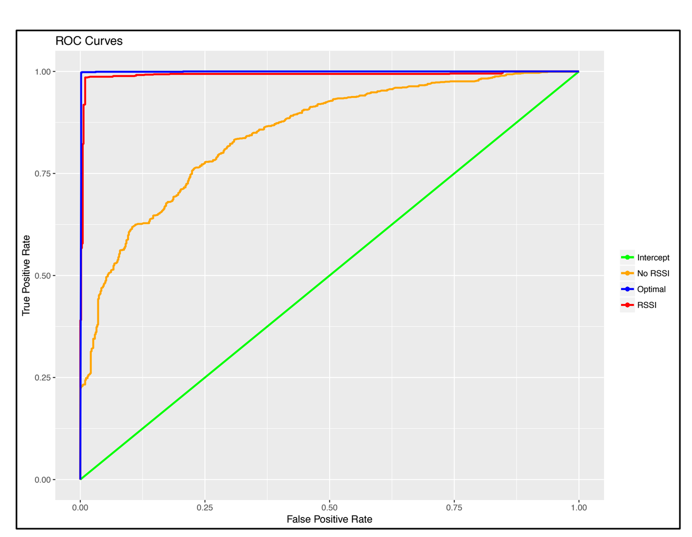
Results:

Number of Trees	Train	CV	Test
1	15.845	25.534	23.940
10	8.824	16.666	16.821
100	7.961	16.021	15.720
200	7.846	15.993	15.631
400	7.769	16.065	15.567
600	7.727	15.930	15.481
800	7.757	15.910	15.515
1000*	7.761	15.688	15.513



Discussion

- RFR (1000 learners) is the most accurate model for predicting RSSI
- SVR (RBF, C = 512, ε = 0.14) outperforms the best OLS model by 20%, validating a nonlinear relationship between RSSI and the feature set
- Adding higher-order terms and interactions marginally improved the prediction accuracy for OLS models
- Feature selection algorithms reduced variance in OLS; however, bias remained problematic with respect to desired performance (17.75 dB)
- Optimal feature sets from regularization and forward search performed worse in SVR and RFR than the baseline feature set of simple regressors
- In order of significance: Network Type, Location, Packet Loss, and RTT



Applications:

- Serves as a personal heuristic for locating areas of optimal and suboptimal cell service
- Benefit cell service providers in determining a favorable deployment of cell towers over a region

Future Work

- Consider learning an Artificial Neural Network (ANN)
- Train a *k*-Nearest Neighbor (kNN) regression model
- Continue exploring the SVR parameter space (may take a very long time)
- Consider using a Weight-Learning Ensemble approach to improve SVR
- Explore RFR performance as a function of max tree depth or max number of nodes

References

[1] Fjelberg, M. "Predicting data traffic in cellular data networks," KTH Royal Institute of Technology, June 2015.

[2] Goudarzi, S. et al. "A Novel RSSI Prediction using Imperialist Competition Algorithm (ICA), Radial Basis Function (RBF) and Firefly Algorithm (FFA) in Wireless Networks," PLOS, July 2016.

[3] Cheng, Y. et al. "High-Precision Wireless Indoor Localization via Weight-Learning Ensemble Support Vector Regression," Research Center for Information Tech. Innovation," Fall 2014.

[4] Mukherjee, T. et al. "RSSI-Based Supervised Learning for Uncooperative Direction-Finding," Intelligent Robotics Inc., Winter 2017.

[5] Cass, P. et al. "Predicting QoE in Cellular Networks using Machine Learning and in-Smartphone Measurements," Austrian Institute of Technology, Fall 2017.