Reconstruct Radio Map with Automatically Constructed Gaussian Process for Localization

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

目录

CONTENT

01

PART ONE

Background

Background

GPS： time consuming

Power consuming

Turn on

meter

Base station

Signal strength

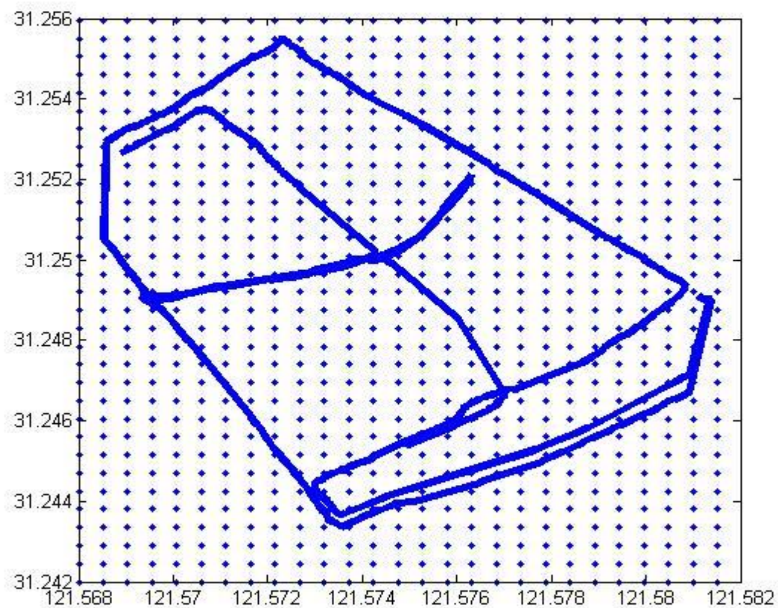
Comparison

- indoor localization use fingerprinting
- creating a radio map
- Received Signal Strength Indicators (RSSI) values obtained from multiple access points (APs)

- a large outdoor environment
- sample thousands of survey sites to construct a fine grain radio map
- a university usually needs hundred thousand training data

Background


轨迹
图表
数据
金杨第一商城
200米
2016-08-30 上午 骑行
未同步
31.67 km
2:35:22
12.23 km/h
0
349 m


31.256
31.254
31.252
31.25
31.248
31.246
31.244
31.242
121.568
121.57
121.572
121.574
121.576
121.578
121.58
121.582

Background

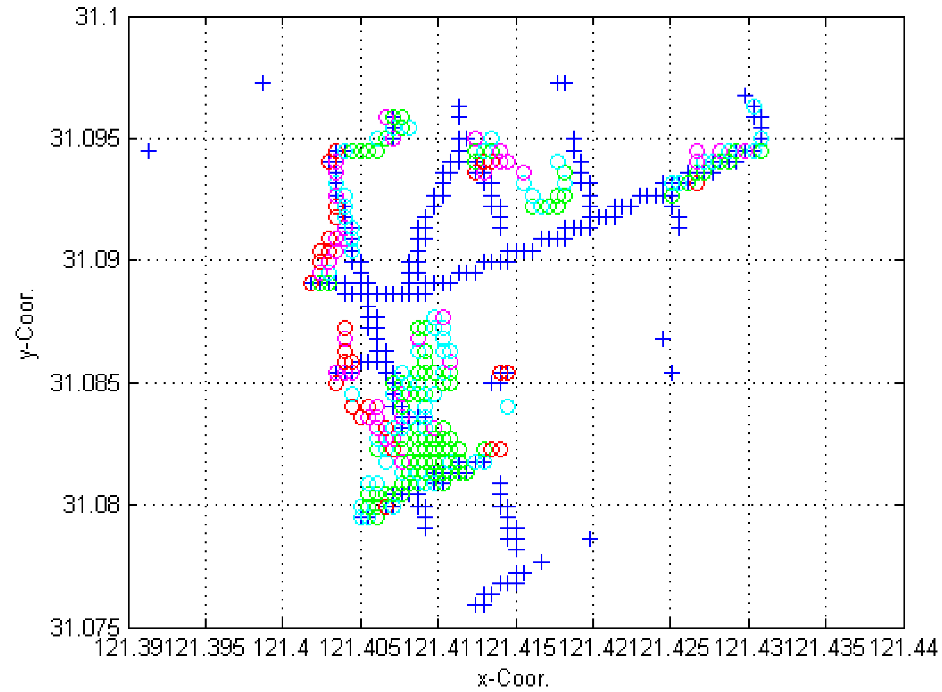
About one million sample data

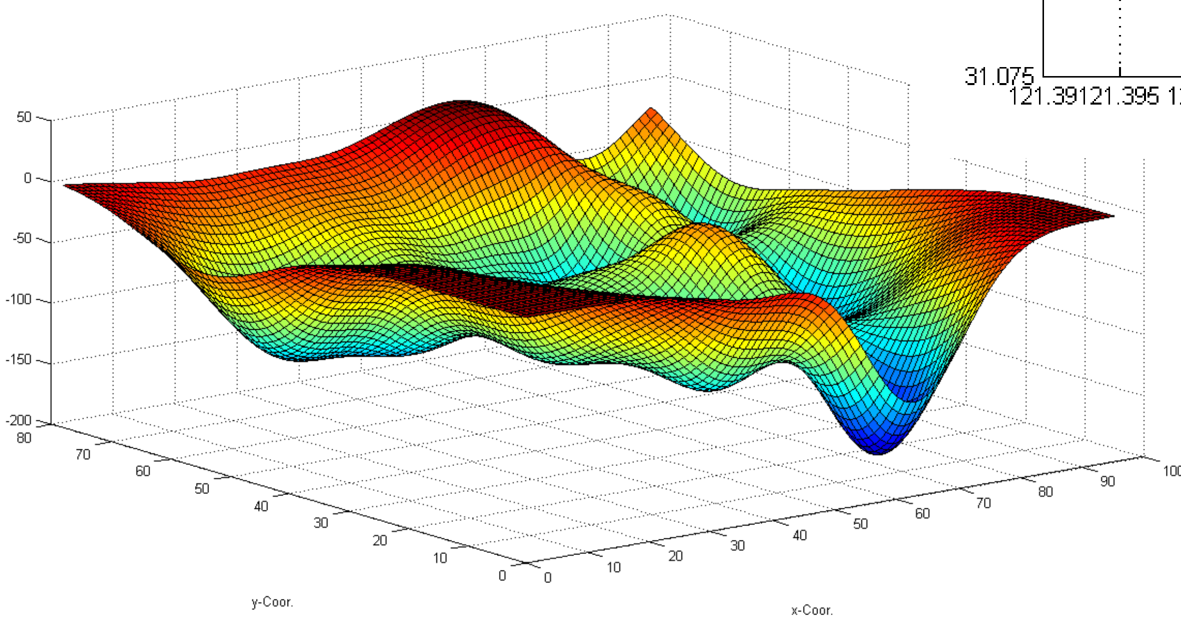
20 square kilometers

SJTU
3G 4G

Yindu Road
3G 4G

Background


31.1
31.095
31.09
31.085
31.08
31.075
y-Coor.
121.39 121.395 121.4 121.405 121.41 121.415 121.42 121.425 121.43 121.435 121.44
x-Coor.


50
0
-50
-100
-150
-200
80 70 60 50 40 30 20 10 0
y-Coor.
0 10 20 30 40 50 60 70 80 90 100
x-Coor.

02

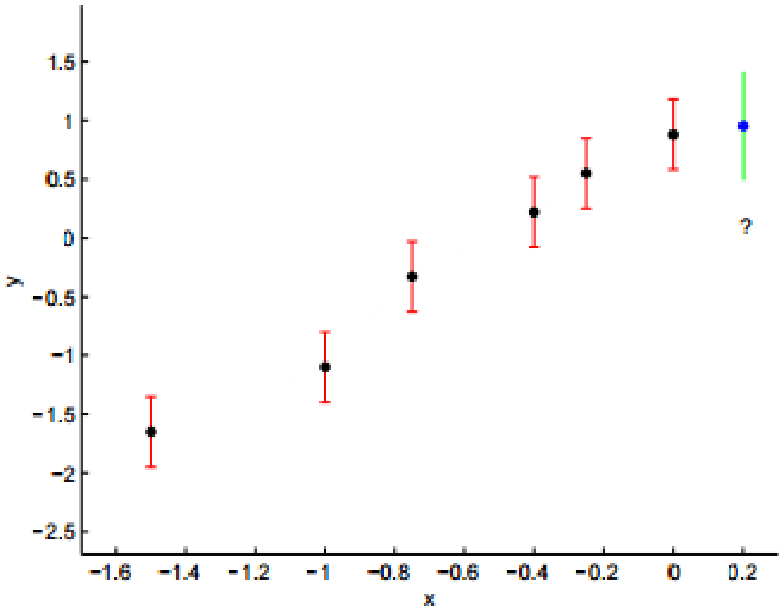
PART TWO

Gaussian Process

GP Motivation

Given a training set $D = \{(\boldsymbol{x}_i, y_i) | i = 1, \ldots, n\}$

How to calculate the output y_* for a new input x_*



Linear regression?

– Least Square Method

Nonlinear regression?

-- Gaussian Process

Relationship to Linear Regression

- In logistic regression, the input to the sigmoid function is $f = \omega^T x + b$ where ω are parameters.
- A Gaussian process places a prior on the space of functions f directly, without parameterizing f.
- Therefore, Gaussian processes are non-parametric
- more general than standard regression the form not limited by a parametric form

Definition

Given a training set $D = \{(\boldsymbol{x}_i, y_i) | i = 1, \dots, n\}$

How to calculate the output y_*(RSS) for a new input x_*（longitude/latitude）

Assume $\boldsymbol{y} = \{y_1, y_2, \dots, y_n\}$ obey multivariate Gaussian Distribution

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \sim N(0, K) \rightarrow \boldsymbol{y} \sim N(0, K)$$

Where

$$K = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \cdots & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & \cdots & k(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_n, x_1) & k(x_n, x_2) & \cdots & k(x_n, x_n) \end{bmatrix}$$

Definition of GP

- A Gaussian process any finite number of which have joint Gaussian distributions.

$$f \sim gp(m, k)$$

- A Gaussian process is fully specified by its mean function m(x) and covariance function k(x,x’).

- Two things to define our GP:
  - choose a form for the mean function.
  - choose a form for the covariance function

GP Regression

For new input data y*
joint distribution defined as:

$$\begin{bmatrix} \mathbf{y} \\ y_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K & K_*^{\mathrm{T}} \\ K_* & K_{**} \end{bmatrix}\right)$$

$$K = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \cdots & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & \cdots & k(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_n, x_1) & k(x_n, x_2) & \cdots & k(x_n, x_n) \end{bmatrix}$$

$$K_* = \begin{bmatrix} k(x_*, x_1) & k(x_*, x_2) & \cdots & k(x_*, x_n) \end{bmatrix} \qquad K_{**} = k(x_*, x_*).$$

高斯过程——均值方差

Get conditional distribution：

$$y_*|\mathbf{y} \sim \mathcal{N}(K_*K^{-1}\mathbf{y},\ K_{**} - K_*K^{-1}K_*^{\mathrm{T}}).$$

Mean and variance:

$$\overline{y}_* = K_*K^{-1}\mathbf{y},$$

$$\mathrm{var}(y_*) = K_{**} - K_*K^{-1}K_*^{\mathrm{T}}.$$

Tune Hyper-Parameters

Hyper-parameters::

$$\boldsymbol{\theta} = \{l, \sigma_f, \sigma_n\} \qquad k(x, x') = \sigma_f^2 \exp\left[\frac{-(x - x')^2}{2l^2}\right] + \sigma_n^2 \delta(x, x')$$

Maxlimum: $p(\boldsymbol{\theta}|\mathbf{x}, \mathbf{y})$

Maxlimum the log likelihood: (conjugate gradients)

$$\log p(\mathbf{y}|X, \boldsymbol{\theta}) = -\frac{1}{2}\mathbf{y}^\top K_y^{-1}\mathbf{y} - \frac{1}{2}\log|K_y| - \frac{n}{2}\log 2\pi.$$

Hyper-parameters

The covariance function defines how smoothly the (latent) function f varies from a given x.

SE kernel: $$k(x, x') = \sigma_f^2 \exp\left[\frac{-(x-x')^2}{2l^2}\right] + \sigma_n^2 \delta(x, x')$$

σ_f^2: overall vertical scale of variation of the latent value.

l: characteristic length-scale

- short means the error bars can grow rapidly away from the data points.
- large implies irrelevant features .

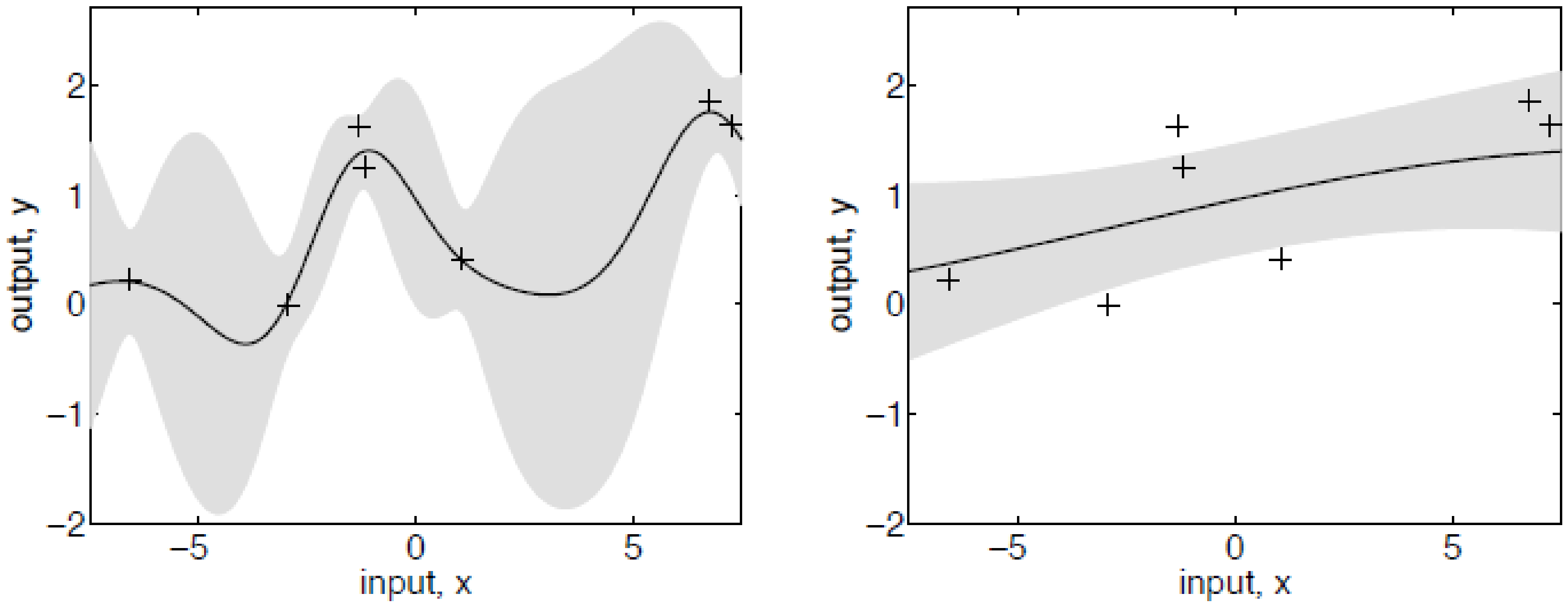
σ_n^2: noise variance

Hyper-parameters

Different hyper parameters

$$\overline{y}_* \pm 1.96\sqrt{\text{var}(y_*)}$$

Bias & Variance trade off



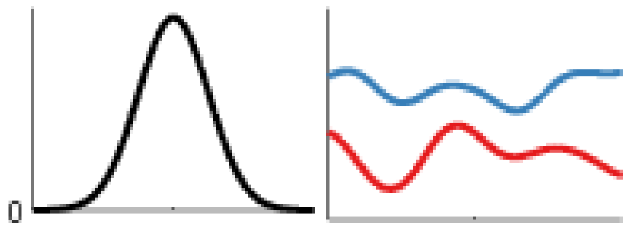
left with small l right with large l

0
3
PART three

Kernel Selection

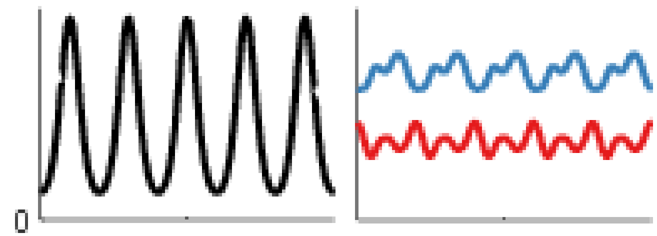
Kernel Selection

Squared Exponential Kernel



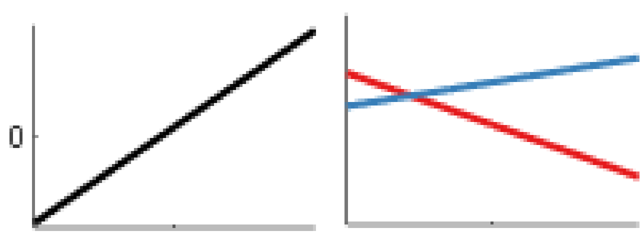
$$k_{\mathrm{SE}}(x, x') = \sigma^2 \exp\left(-\frac{(x-x')^2}{2\ell^2}\right)$$

Periodic Kernel



$$k_{\mathrm{Per}}(x, x') = \sigma^2 \exp\left(-\frac{2\sin^2(\pi|x-x'|/p)}{\ell^2}\right)$$

Linear Kernel

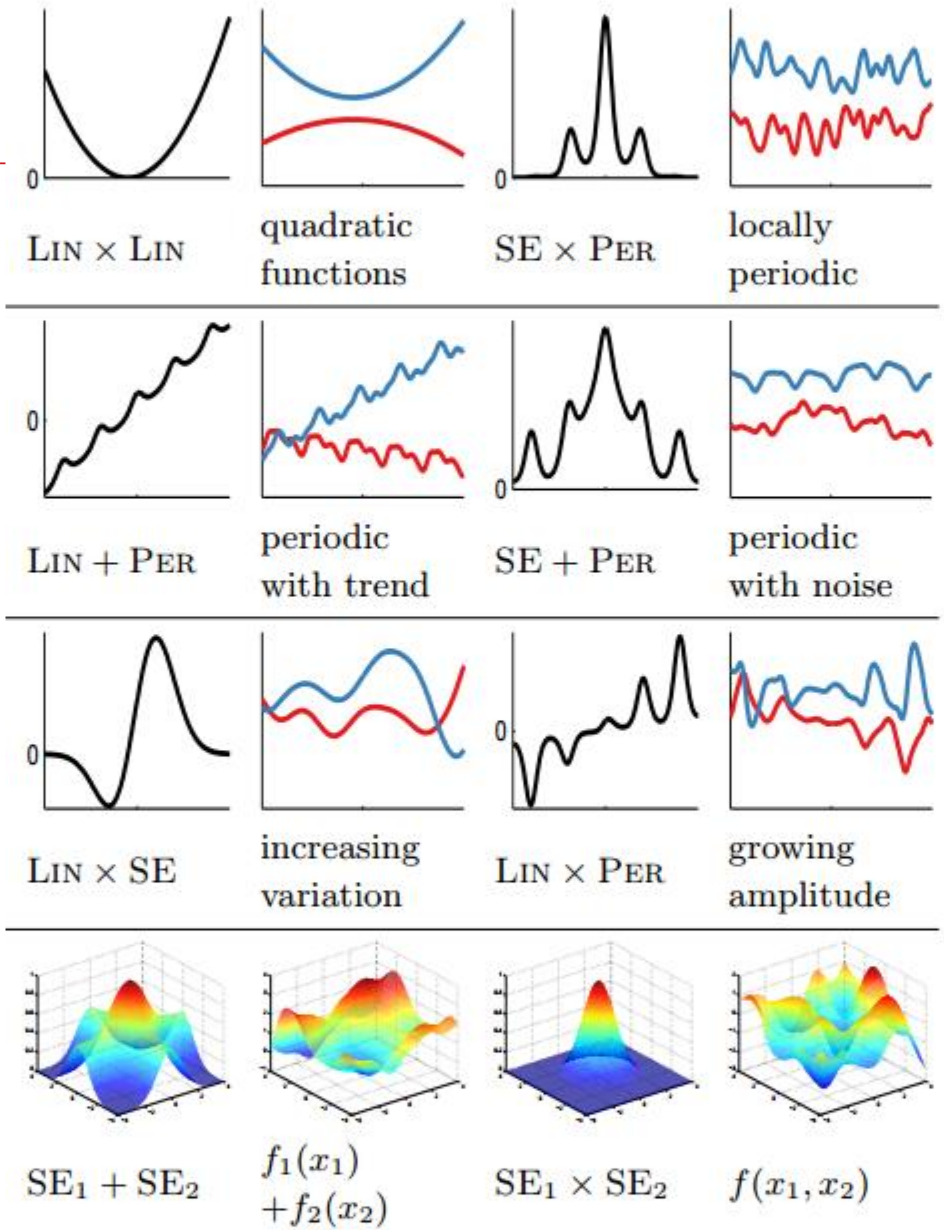


$$k_{\mathrm{Lin}}(x, x') = \sigma_b^2 + \sigma_v^2(x-c)(x'-c)$$

Composing Kernel

（1） ka + kb = ka(x, x’) + kb(x, x’)（Summation）

（2） ka × kb = ka(x, x’) × kb(x, x’)（Multiplication）



Automatic Model Construction

choosing the structural form of the kernel: a black art

$$k_{\mathrm{SE}}(x,x') = \sigma^2 \exp\left(-\frac{(x-x')^2}{2\ell^2}\right)$$

$$k_{\mathrm{PER}}(x,x') = \sigma^2 \exp\left(-\frac{2\sin^2(\pi(x-x')/p)}{\ell^2}\right)$$

$$k_{\mathrm{LIN}}(x,x') = \sigma_b^2 + \sigma_v^2(x-\ell)(x'-\ell)$$

$$k_{\mathrm{RQ}}(x,x') = \sigma^2 \left(1+\frac{(x-x')^2}{2\alpha\ell^2}\right)^{-\alpha}$$

$$\mathrm{BIC}(M) = -2\log p(D \mid M) + |M| \log n$$

BIC trades off model fit and complexity

- Search over sums and products of kernels
- Maximizing the BIC(M)
- Show how any model be decomposed into different parts

04

PART FOUR

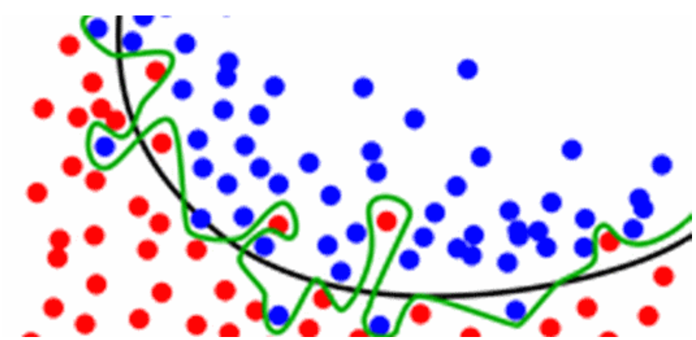
Model Ensemble

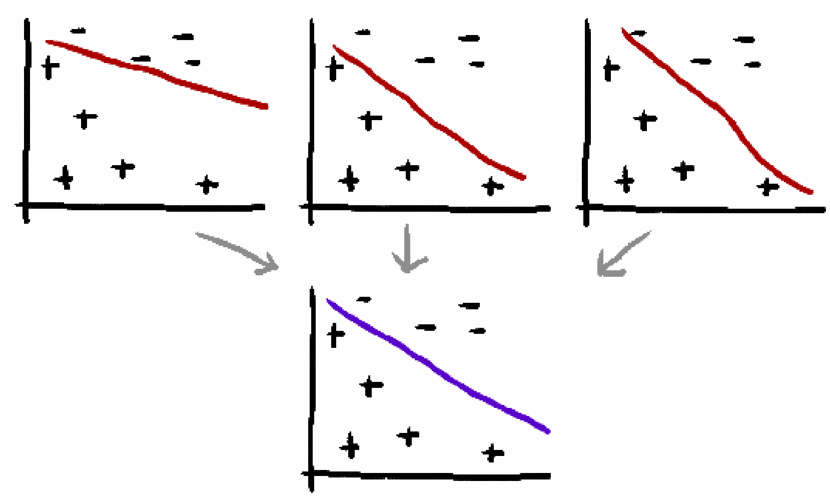
Model Ensemble

Don’ t Overfit!

Averaging multiple different green lines should bring us closer to the black line.

1)average out biases
2)reduce the variance
3)unlikely to overfit





Model Ensemble

Linear regression

SVR

Gradient Tree Boosting

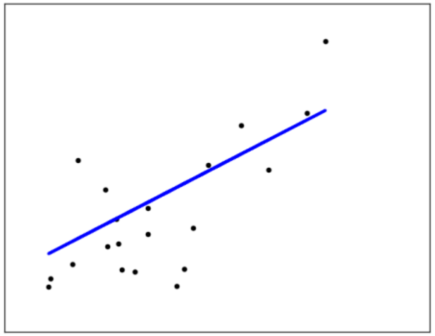
Xgboost

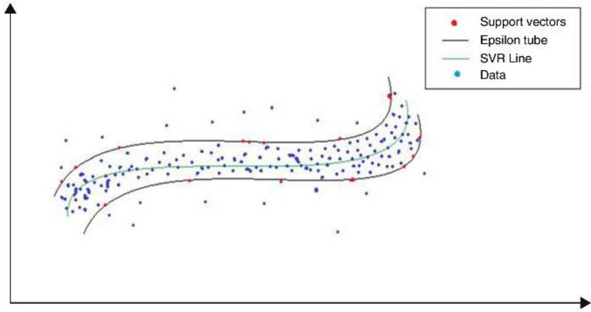
GP (RQ)

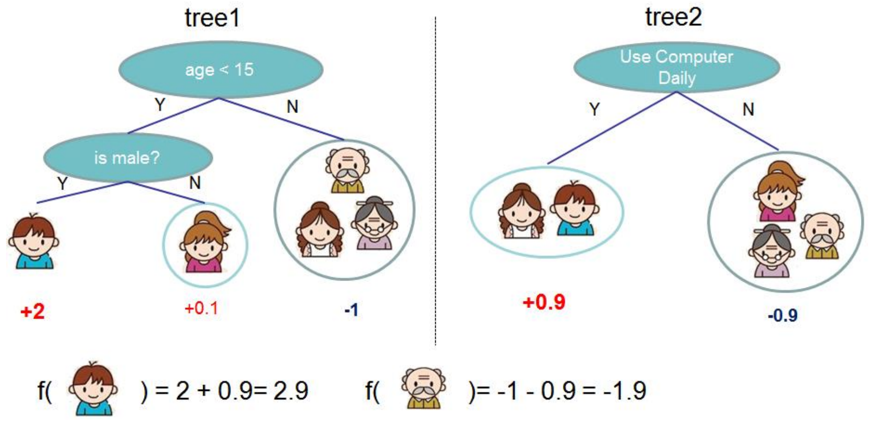
GP (Compose)

Rate Averaging

Blending





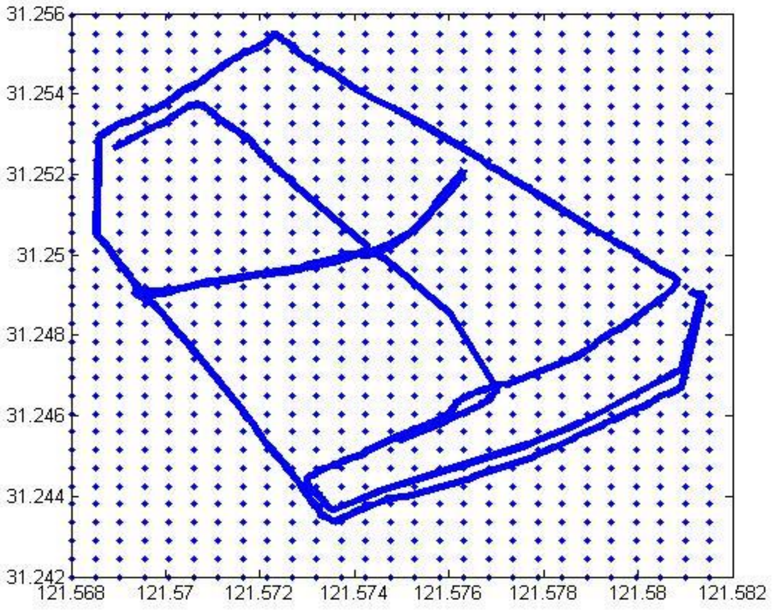


0 5

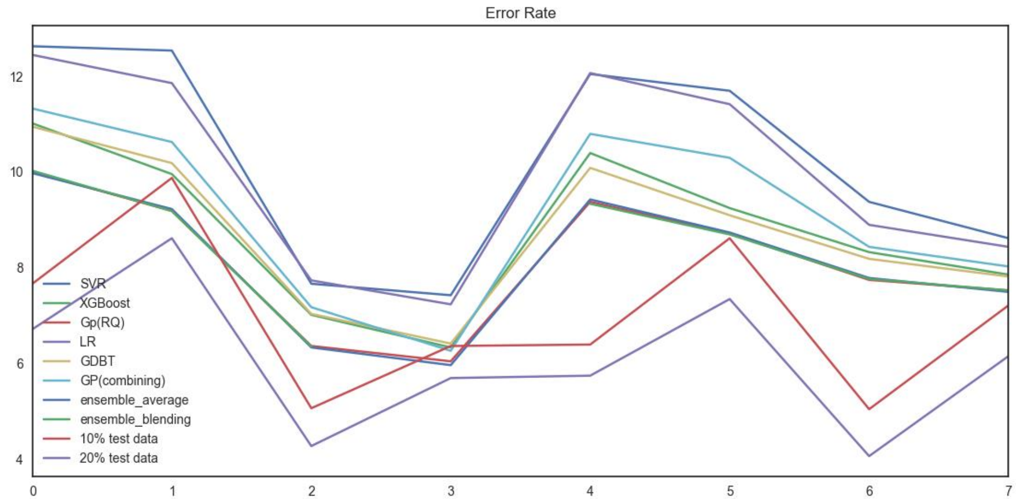
PART FIVE

Experiment Results

Experiment Results



Experiment Results



Unnamed: 0	SVR	XGBoost	Gp(RQ)	LR	GDBT	GP(combining)	nsemble_averag	ısemble_blend	10% small road	20% small road
SJTU_3G…	12.6	11	10	12.4	10.9	11.3	9.98	10	7.67	6.72
SJTU_3G…	12.5	9.96	9.19	11.9	10.2	10.6	9.23	9.19	9.88	8.62
SJTU_4G…	7.67	7.02	6.37	7.74	7.04	7.18	6.34	6.36	5.07	4.28
SJTU_4G…	7.43	6.34	6.05	7.24	6.42	6.27	5.97	nan	6.37	5.7
YINDU_3…	12.1	10.4	9.38	12.1	10.1	10.8	9.43	9.34	6.4	5.75
YINDU_3…	11.7	9.25	8.72	11.4	9.1	10.3	8.74	8.7	8.62	7.35
YINDU_4…	9.38	8.33	7.75	8.9	8.19	8.44	7.79	7.77	5.05	4.07
YINDU_4…	8.62	7.86	7.53	8.44	7.82	8.03	7.5	7.53	7.22	6.16

06

PART SIX

Q&A