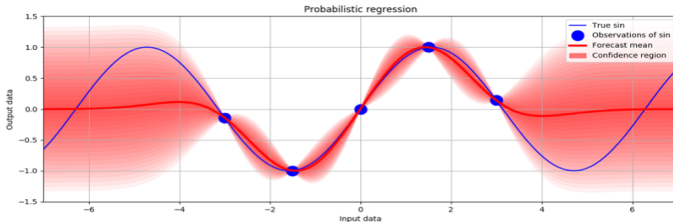


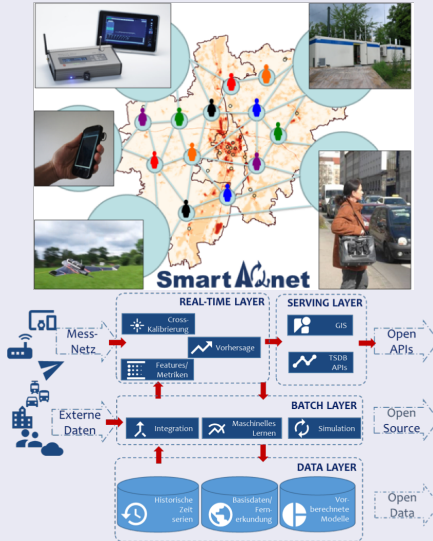
Data-driven Modeling and Optimization in Fluid Mechanics, KIT, 2019

Gaussian process regression for heterogeneous measuring networks of environmental data.

Dr. rer. nat. Johannes Riesterer - KIT/TECO



Heterogeneous measuring network



SmartAQNet (api.smartaq.net). Measuring network in Augsburg.

Total Number of Observations

12585217

Active Sensors (last 24h)

20

Map showing active Sensors



← Back to Thing: Crowdsensing Node (SDS011)

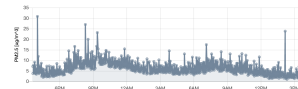
PROPERTIES THING SENSOR OBSERVATIONS OBSERVEDPROPERTY

2019-09-11 15:35:40 - 2019-09-12 15:34:38 -

Show Live Data

RESULTS

COMPARE TO ...



Result Time	Result [µg/m³]
2019-09-12 15:34:38	5.00
2019-09-12 15:34:05	4.65
2019-09-12 15:33:32	2.70
2019-09-12 15:32:58	8.28
2019-09-12 15:32:24	6.78
2019-09-12 15:31:51	3.35

Technologies



Smart Data
Innovation Lab



SmartAQNet



HelmholtzZentrum münchen

Deutsches Forschungszentrum für Gesundheit und Umwelt

Automated Quality Assessment of (Citizen) Weather Stations; Julian Bruns, Johannes Riesterer, Bowen Wang, Till Riedel, Michael Beigl; GI-Forum (2018).

Johannes Riesterer, Sebastian Lerch, Matthias Budde, Julian Bruns, Stanislav Arnaudov, Till Riedel, Michael Beigl (2019) Stochastische Regressionsmodelle zur Verbesserung der Datenqualität, Kalibrierung und Interpolation von Umwelt-und Luftdaten in verteilten Messnetzen aus Low-Cost Sensoren, Umwelteinflüsse erfassen, simulieren, bewerten - 48. Jahrestagung der GUS 2019

Stefan Hinterreiter, Matthias Budde, Klaus Schäfer, Johannes Riesterer, Till Riedel, Marcel Köpke, Josef Cyrus, Stefan Emeis, Thomas Gratza, Marcus Hank, Andreas Philipp, Erik Petersen, Johanna Redelstein, Jürgen Schnelle-Kreis, Duick Young, Michal Kowalski, Volker Ziegler, Michael Beigl (2018) SmartAQnet – Neuer smarterer Weg zur räumlichen Erfassung von Feinstaub, AGIT – Journal für Angewandte Geoinformatik 4.

Gaussian Processes for Machine Learning Carl Edward Rasmussen and Christopher K. I. Williams The MIT Press, 2006. Online: <http://www.gaussianprocess.org/gpml/>

The Nature of Statistical Learning Theory (Information Science and Statistics) Vapnik, Vladimir

Assumptions

Let $Y : \mathbb{R}^n \rightarrow \mathbb{R}$, $X : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be random variables.

Problem

For a given loss function L , find function $f : \mathbb{R}^m \rightarrow \mathbb{R}$ with $f = \operatorname{argmin}_h L(Y, h(X))$.

Specific solution

$$L(Y, h(X)) := \int (Y - h(X))^2 d\mu$$
$$\Rightarrow f = \mathbb{E}(Y|X) = \int Y \cdot p(Y|X) d\mu$$

Assumptions

Let $D = \{(y_i, x_i)_i\}$ be samples (Big Data) of the random variables $Y : \mathbb{R}^n \rightarrow \mathbb{R}$, $X : \mathbb{R}^n \rightarrow \mathbb{R}^m$ with unknown distributions .

Algorithmic problem

Choose a model $Y = f(X) = f_\theta(X)$ where $f_\theta(X)$ is a parametric family. Find optimal parameters

$$f = \operatorname{argmin}_{\{f_\theta \mid \theta \in \Theta\}} L(Y, f_\theta(X))$$

$$L(Y, h(X)) \doteq \sum_D (y - h(x))^2$$

$$D = D_{\text{Train}} \cup D_{\text{Test}} \cup D_{\text{Validation}}$$

Solution

Compute "posterior predictive distribution" $p(y^*|x^*, D)$

Definition

$\mathbb{E}(y^*|x^*, D)$ is called prediction for the feature x^* and $\mathbb{V}(y^*|x^*, D)$ is a measure of the accuracy of the prediction.

Bayesian inference

$$p(y^*|x^*, D) = \int_{\Theta} p(y^*, \theta|x^*, D) d\theta = \int_{\Theta} \underbrace{p(y^*|\theta, x^*, D)}_{=p(y^*|\theta, x^*)} \cdot p(\theta|D) d\theta$$

Bayes Theorem

$$\text{posterior} = \frac{\text{likelihood} \cdot \text{prior}}{\text{marginal likelihood}}$$
$$p(\theta|D) = \frac{p(D|\theta) \cdot p(\theta)}{p(D)}$$

Conjugate priors

Analytically solvable for conjugate priors.

Markov-Chain-Monte-Carlo

Or solve integrals via sampling:

$$\sum_{x_i} \frac{g(x_i)}{p(x_i)} \rightarrow \int_{\Omega} g \, d\mu$$

Linear Bayesian Regression

Model

$Y = f(X) = X^t \cdot \omega + \epsilon$ with weights $\omega \sim \mathcal{N}(0, \Sigma)$, $\epsilon \sim \mathcal{N}(0, \sigma^2)$ and X constant.

Posterior predictive distribution

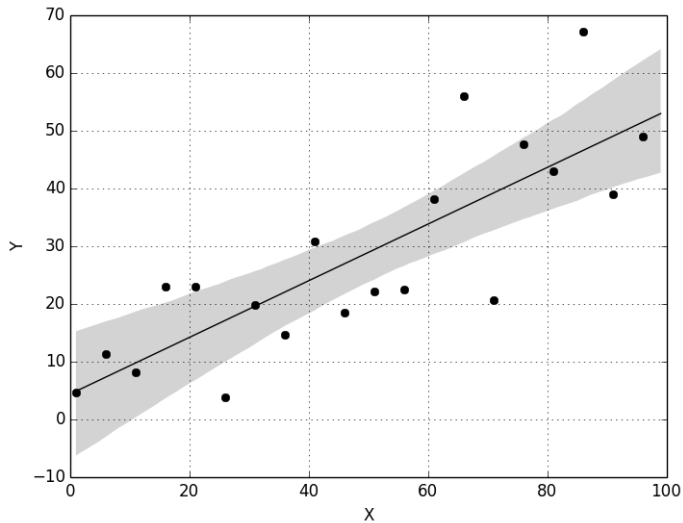
$$y^* | x^*, D \sim \mathcal{N} \left(\frac{1}{\sigma^2} (x^*)^t A^{-1} X^t Y, (x^*)^t A^{-1} x^* \right)$$

with $A = \frac{1}{\sigma^2} X^t X + \Sigma^{-1}$.

Prediction

$$\begin{aligned} \mathbb{E}(y^* | x^*, D) &= \frac{1}{\sigma^2} (x^*)^t A^{-1} X^t Y \\ \mathbb{V}(y^* | x^*, D) &= (x^*)^t A^{-1} x^* \end{aligned}$$

Linear Bayesian Regression

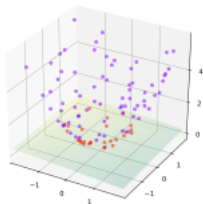
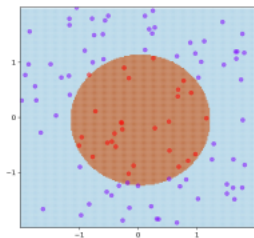


Kernel Trick

$$f(\mathbf{X}) = \phi(\mathbf{X})^t \cdot \omega \text{ mit } \phi : \mathbb{R}^n \rightarrow \mathbb{R}^N$$

Example

$$\phi(\mathbf{X})^t = (x_1, x_2, x_1^2 + x_2^2)$$



Bayesian Regression

Similar computation as in the case of linearer Bayesian Regression:

$$y^*|x^*, D \sim \mathcal{N} \left(\frac{1}{\sigma^2} \underbrace{\phi(x^*)^t A^{-1} \phi(x)}_{:=k(x^*, x)} Y, \phi(x^*)^t A^{-1} \phi(x^*) \right)$$

with $A = \frac{1}{\sigma^2} \phi(x)^t \phi(x) + \Sigma^{-1}$

Kernel

A function of the form

$$k(x^*, x) = \langle \phi(x^*), \phi(x) \rangle$$

is called kernel.

Mercers Theorem

Let $k(s, t) = k(t, s)$ and $\int_{X \times X} k(s, t) f(s) f(t) > 0$ for all $f \in L^2(X)$, then

$$k(s, t) = \sum_{i=1}^{\infty} \lambda_i \phi_i(s) \phi_i(t)$$

where ϕ_i are eigenfunctions corresponding to the eigenvalues λ_i of the linear operators $T_k f := \int_X k(\cdot, t) f(t)$.

Stochastic Prozess

A stochastic process is a family of random variables $\{f_x | x \in \mathcal{X}\}$.

Gaussian process

A stochastic process is called Gaussian process

$f_x \sim \mathcal{GP}(m(x), k(x, x'))$, if

$$\begin{pmatrix} f_{x_1} \\ \vdots \\ f_{x_n} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} m(x_1) \\ \vdots \\ m(x_n) \end{pmatrix}, \begin{pmatrix} k(x_1, x_1) & \dots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \dots & k(x_n, x_n) \end{pmatrix} \right)$$

for all finite subsets $X = (x_1, \dots, x_n) \in \mathcal{X}$. $k(x, x')$ is called kernel. They have to be chosen such that the matrix is positive definite and symmetric for all finite subsets X .

Example

$f_x = \phi(x)^t \cdot \omega$ with $\omega \sim \mathcal{N}(0, \Sigma)$ is a $\mathcal{GP}(m(x), k(x, x'))$ with

$$\mathbb{E}(f_x) = \phi(x)^t \mathbb{E}(\omega) = \underbrace{0}_{=: m(x)}$$

$$\mathbb{E}(f_x f_{x'}) = \phi(x)^t \mathbb{E}(\omega \omega^t) \phi(x') = \underbrace{\phi(x)^t \Sigma \phi(x')}_{:= k(x, x')}$$

Prior distribution

Let $f \sim \mathcal{GP}(0, k(x, x'))$ be a Gaussian process and let $f = (f_{x_1} \dots f_{x_n})$ be known for $X = (x_1, \dots, x_n)$. Thus

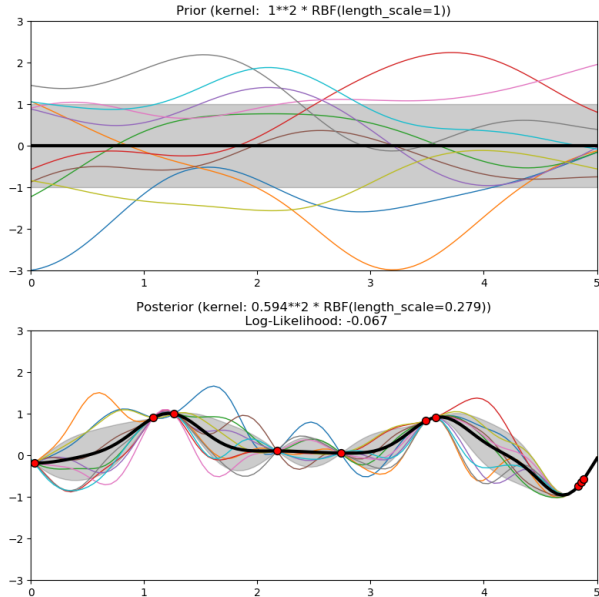
$$\begin{pmatrix} f_X \\ f_{X^*} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} K(X, X) & K(X, X^*) \\ K(X^*, X) & K(X^*, X^*) \end{pmatrix} \right)$$

for arbitrary points $X^* = (x_1^*, \dots, x_n^*)$

Posterior predictive distribution

$$\begin{aligned} \Rightarrow \quad f_{X^*} | X^*, f_X, X &\sim \mathcal{N}(\mu, \Sigma) \\ \mu &:= K(X^*, X)K(X, X)^{-1} \cdot f \\ \Sigma &:= K(X^*, X^*) - K(X^*, X)K(X, X)^{-1}K(X, X^*) \end{aligned}$$

Gaussian process regression



Prediction

For a prior \tilde{f} we define the prediction by

$$R(\tilde{f}) := \mathbb{E}((\tilde{f}|\tilde{X}, X, f)) = K(\tilde{X}, X)K(X, X)^{-1} \cdot f$$

The variance

$$\mathbb{V}((\tilde{f}|\tilde{X}, X, f)) = K(\tilde{X}, \tilde{X}) - K(\tilde{X}, X)K(X, X)^{-1}K(X, \tilde{X})$$

is a measure of the accuracy of the prediction.

Kernel

Kernel	Funktion
konstant	σ_0^2
linear	$\sum_{d=1}^D \sigma_d^2 x_d x'_d$
polynomial	$(x \cdot x' + \sigma_0^2)^p$
squared exponential	$\exp\left(-\frac{r^2}{2l^2}\right)$
Matérn	$\frac{1}{2^{\nu-1} \Gamma(\nu)} \left(\frac{\sqrt{2\nu} r}{l}\right)^\nu K_\nu\left(\frac{\sqrt{2\nu} r}{l}\right)$
exponentiell	$\exp\left(-\frac{r}{l}\right)$
γ -exponentiell	$\exp\left(-\left(\frac{r}{l}\right)^\gamma\right)$
rational quadratisch	$\left(1 + \frac{r^2}{2\alpha l^2}\right)^{-\alpha}$

Numerical aspects

One has to invert a large matrix, which is a non trivial problem.

Common solutions

Use divide and conquer algorithms.

Use subsets for computation.

Solve posterior predictive distribution with sampling method.

Quality model

Let $\{(p_i, y_i)\}$ be a set of sensors at position p_i and measuring y_i of unknown quality. Assign each sensor a quality value $q_i \in (0, \infty)$.

Kernel

$$K((p_i, q_i), (p_j, q_j)) := K_{Matern}(p_i, p_j) + K_q(q_i, q_j)$$

$$K_q(q_i, q_j) := \begin{cases} \frac{1}{q_i^2} & i=j \\ 0 & \text{else} \end{cases}$$

Model with hyperparameter

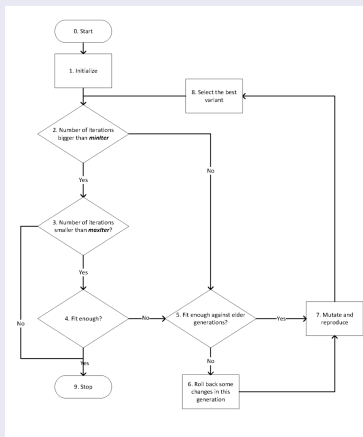
Model

We get a gaussian process regression model $GP(Q)$ with hyperparameters $Q = \{(q_i)\}$

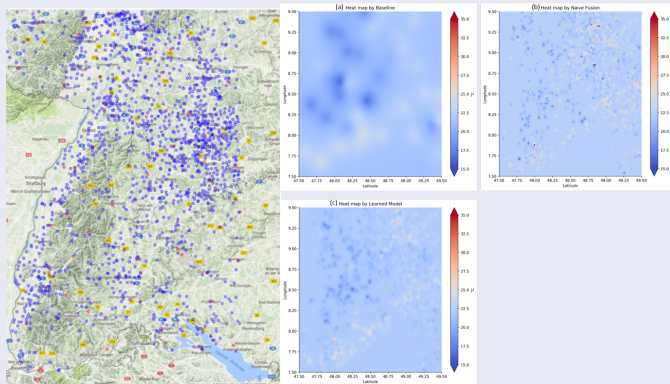
Optimierung

Optimise $GP(Q)$ with respect of Q . Use genetic algorithm for example.

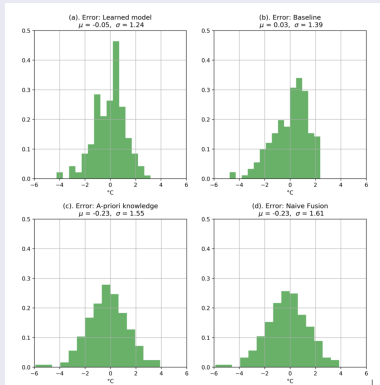
Flowchart



Error analysis



Error analysis



Significant with $\alpha = 0.05$

Master thesis with Hiwi-Job

Stochastic regression models for heterogeneous measuring networks.

In cooperation with Prof. Dr. Grothe - KIT/IOR and Dr. Sebastian Lerch - KIT/CST.

Hiwi-Job

Topological data analysis for heterogeneous measuring networks.

In cooperation with Prof. Dr. Sauer - KIT/IAG and Dr. Schrödl-Baumann -KIT/IAG.

Application

riesterer@teco.edu

More informations: <http://www.teco.edu/people/riesterer/>