Learning from Multiple Annotators with Gaussian Processes

Perry Groot[1,2,*], Adriana Birlutiu[2], Tom Heskes[2]

[1] Technical University Eindhoven - Dept. of Electrical Engineering - Control Systems
Potentiaal 4.28, 5600 MB Eindhoven - the Netherlands
[2] Radboud University Nijmegen - Intelligent Systems
Heyendaalseweg 135, 6525 AJ Nijmegen - the Netherlands
[*] Corresponding author, e-mail: pcgroot@gmail.com

1. Introduction

In many supervised learning tasks it can be costly or infeasible to obtain objective, reliable labels from experts. We may, however, be able to obtain a large number of subjective, possibly noisy, labels from multiple annotators. Typically, annotators have different levels of expertise (i.e., novice, expert) and there is considerable disagreement among annotators.

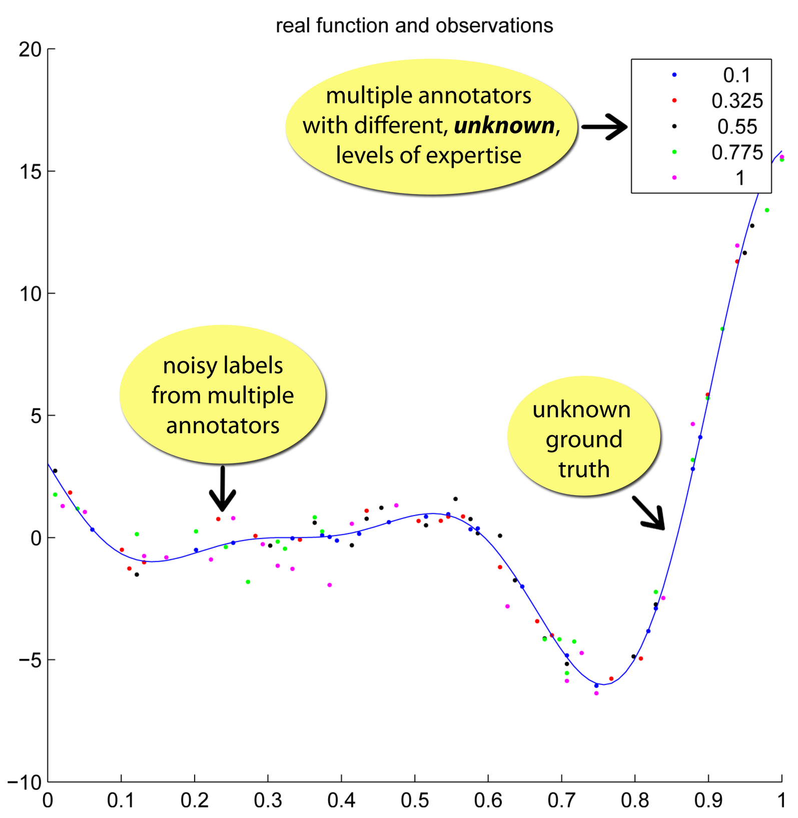

Figure 1: *Regression with labels from multiple noisy annotators.*

In recent years there has appeared a lot of interest in multi-annotator problems without an absolute gold standard. Previous work, however, has focused mainly on parametric models [1, 3, 4]. We present a flexible non-parametric approach based on Gaussian processes for regression with multiple labels.

2. Gaussian Process Regression

Let $\mathcal{D}$ be a data set of N observed D-dimensional input vectors $\boldsymbol{X} = \{\boldsymbol{x}_n\}_{n=1}^N$ and corresponding real-valued outputs $\boldsymbol{Y} = \{y_n\}_{n=1}^N$. We assume that outputs follow from a latent function f that are corrupted by zero mean Gaussian noise, i.e., $y = f(\boldsymbol{x}) + \epsilon$ with $\epsilon \sim \mathcal{N}(0, \sigma^2)$. A **Gaussian process** defines a prior distribution over functions f giving a multi-variate Gaussian distribution on any finite subset of latent variables, i.e., the function values $f(\boldsymbol{x})$. The Gaussian process is completely specified by a mean function (which we assume zero without loss) and a covariance function. In particular $p(f|\boldsymbol{X}) = \mathcal{N}(f|\boldsymbol{0}, \boldsymbol{K}_{NN})$ with $K_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$ for which we used the well-known Gaussian (or squared-exponential) covariance function. For more details see [2].

3. Multi-Annotator Regression

We assume that observations are obtained from M annotators, each providing noisy labels that adhere to a Gaussian distribution $\mathcal{N}(0, \sigma_m^2)$ where σ_m^2 represents the *unknown* noise level of the m-th annotator. Let

$$\frac{1}{\hat{\sigma}_i^2} = \sum_{m\sim i} \frac{1}{\sigma_m^2}, \qquad \hat{y}_i = \hat{\sigma}_i^2 \sum_{m\sim i} \frac{y_i^m}{\sigma_m^2}, \qquad \hat{\boldsymbol{\Sigma}} = \mathrm{diag}(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_I^2), \tag{1}$$

with $m \sim i$ denoting the sum over annotators m that annotated sample $\boldsymbol{x}_i$. It can be shown that the multi-annotator predictive distribution is given by:

$$\begin{aligned} \overline{f}_{\mathcal{D}}(\boldsymbol{x}) &= k(\boldsymbol{x}, \boldsymbol{X})(\boldsymbol{K} + \hat{\boldsymbol{\Sigma}})^{-1}\hat{\boldsymbol{Y}}, \\ \mathrm{cov}_{\mathcal{D}}(f(\boldsymbol{x}), f(\boldsymbol{x}')) &= k(\boldsymbol{x}, \boldsymbol{x}') - k(\boldsymbol{x}, \boldsymbol{X})(\boldsymbol{K} + \hat{\boldsymbol{\Sigma}})^{-1} k(\boldsymbol{X}, \boldsymbol{x}'), \end{aligned} \tag{2}$$

which closely follows the single-annotator model, but uses a weighted output $\hat{\boldsymbol{Y}}$ and covariance $\hat{\boldsymbol{\Sigma}}$ that is no longer homogeneous as it depends on the data sample. Hyperparameters can be optimized automatically by minimizing the negative log marginal likelihood:

$$\begin{aligned} -\log p(\boldsymbol{Y}) = \frac{1}{2}\log|\boldsymbol{K} + \hat{\boldsymbol{\Sigma}}| + \frac{1}{2}\hat{\boldsymbol{Y}}^T(\boldsymbol{K} + \hat{\boldsymbol{\Sigma}})^{-1}\hat{\boldsymbol{Y}} + \frac{N}{2}\log(2\pi) + \\ -\frac{1}{2}\log|\hat{\boldsymbol{\Sigma}}| - \sum_i \sum_{m\sim i} \log\frac{1}{\sigma_m} + \frac{1}{2}\sum_i \sum_{m\sim i} \frac{(y_i^m)^2}{\sigma_m^2} - \frac{1}{2}\sum_i \frac{\hat{y}_i^2}{\hat{\sigma}_i^2}. \end{aligned} \tag{3}$$

4. Experiments

We validated the Gaussian process multi-annotator model on the 'Boston housing' dataset by randomly annotating a portion of the data set using three annotators with noise levels of 0.25, 0.5, and 0.75. We report the root mean squared error ($RMSE(x, y) = \sqrt{\frac{1}{N}\sum_{i=1}^N (x_i - y_i)^2}$) for both the prediction of the targets and the hyperparameter prediction of the annotator noise-levels.

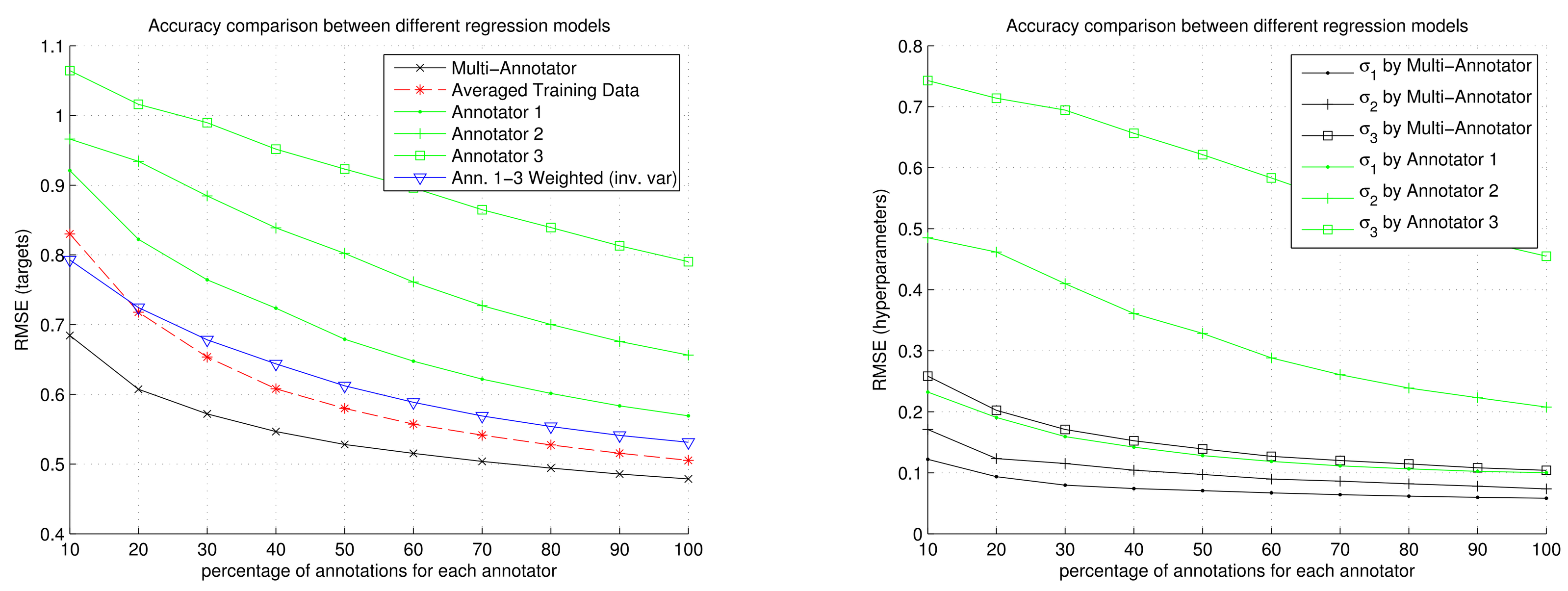

Figure 2: *The RMSE of the Gaussian process multi-annotator model, the Gaussian process fitted to the average response, and Gaussian processes fitted to each individual annotator on the 'Boston housing' dataset (506 instances, 13 features). Left: RMSE for predicted targets. Right: RMSE for predicted noise-level hyperparameters.*

5. Conclusions

The Gaussian process framework provides a principled *non-parametric* framework that can automatically estimate the reliability of individual annotators from data without the need of prior knowledge. Experimental results show that the proposed Gaussian process multi-annotator model outperforms models that either average the training data or weigh individually learned single-annotator models.

References

[1] Ristovski, K., Das, D., Ouzienko, V., Guo, Y., Obradovic, Z.: Regression Learning with Multiple Noisy Oracles. In: 19th European Conference on Artificial Intelligence, pp. 445-450 (2010)

[2] Rasmussen, C.E., Williams, C.K.I.: Gaussian processes for machine learning. MIT Press, Cambridge, MA (2006)

[3] Raykar, V.C., Yu, S., Zhao, L.H., Jerebko, A., Florin, C.: Supervised learning from multiple experts: whom to trust when everyone lies a bit. In: Proceedings of the 26th Annual International Conference on Machine Learning, pp. 889-896 (2009)

[4] Raykar, V.C., Zhao, L.H., Valadez, G.H., Florin, C., Bogoni, L., Moy, L.: Learning from crowds. JMLR, 11, 1297-1322 (2010)

Acknowledgement: This work has been carried out as part of the OCTOPUS project with Océ Technologies B.V. under the responsibility of the Embedded Systems Institute. This project is partially supported by the Netherlands Ministry of Economic Affairs under the Bsik program.