Assignment V: Gaussian Process Regression

Exercises in Machine Learning (190.013), SS2022

Stefan Nehl[1]

[1]*stefan-christopher.nehl@stud.unileoben.ac.at, MNr: 00935188*, Montanuniversität Leoben, Austria

May 26, 2022

In the fifth assignment, I had to describe the Guassian Process Regression and implement it with the library GPy. Furthermore, I had to test different kernel implementations and hyper parameters and verify and compare the results with the results of the last assignment.

1 Gaussian Process Regression

Gaussian processes are a class of nonparametric models for machine learning. They are commonly used for modeling spatial and time series data. (Powell, 2021) The *Gaussian Process Regression* uses the *Multivariate Conditional Distribution*.

$$f(\boldsymbol{x'}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^{\mathrm{k}}|\boldsymbol{\Sigma}|}} exp^{-\frac{1}{2}(\boldsymbol{x'} - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x'} - \boldsymbol{\mu})}$$

With $\boldsymbol{x'} = \{x'_1, ..., x'_k\}$, $\boldsymbol{\Sigma}$ the covariance matrix and $|\boldsymbol{\Sigma}|$ the determinant of the covariance matrix. The covariance is determined by the covariance function of the kernel and has to be positive definite. (Rueckert, 2022) I tried three different kernels for the *Gaussian Process Regression*. The *Linear Kernel*, the *RBF, Radial Basis Function* and the *Matern 52*.

1.1 Linear Kernel

The *Linear Kernel* is based on linear classification. This classification is based on the linear combination of the characteristics. The decision function can be described with:

$$d(\boldsymbol{x}) = \boldsymbol{w}^T\phi(\boldsymbol{x}) + b$$

where $\boldsymbol{w}$ is the weight vector, b a biased value and $\phi(\boldsymbol{x})$ a higher dimensional vector of $\mathbf{x}$. If $\boldsymbol{w}$ is a linear combination of training data $\boldsymbol{w}$ can be calculated with:

$$\boldsymbol{w} = \sum_{i=1}^{l} \boldsymbol{\alpha}_i\phi(\boldsymbol{x_i})$$

for some $\boldsymbol{\alpha} \in \mathbf{R}^1$ The kernel function can be calculated with

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \phi(\boldsymbol{x}_i))^T\phi(\boldsymbol{x}_j))$$

(Guo-Xun Yuan and Lin, 2021) The linear kernel is good for data with a lot of features. That's because mapping the data to a higher dimensional space does not really improve the performance. (KOWALCZYK, 2014) The implementation in the *GPy* library was the following:

$$K(x, y) = \sum_{i=1}^{D} \sigma_i^2 x_i y_i$$

Where D defines the dimension and σ_i^2 the variance for each dimension.

1.2 RBF

The *Radial Basis Function* is one of the most used kernels. It's similar to the *Gaussian distribution*. The kernel calculates the similarity or how close two points are to each other.

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \ epx(-\frac{||\boldsymbol{x}_i - \boldsymbol{x}_j||}{2\sigma^2})$$

Where $||\boldsymbol{x}_i - \boldsymbol{x}_j||$ is the euclidean (L_2-Norm) and σ the variance and the hyper parameter. (Sreenivasa, 2020)

$$K(r) = \sigma^2 \exp\left(-\frac{1}{2}r^2\right)$$

Implementation in the *GPy* library where $r = |\boldsymbol{x}_i - \boldsymbol{x}_j|$ and σ^2 the variance.

1.3 Matern 52

The *Matern 52* is a generalization of the *RBF* kernel.

$$K(r) = (1 + \frac{\sqrt{5}r}{l} + \frac{5r^2}{3l^2})\ exp(-\frac{\sqrt{5}r}{l})$$

Where $r = |\boldsymbol{x}_i - \boldsymbol{x}_j|$ and l a positive parameter.(Williams, 2006) However, if I take a look in the code of the *GPy* implementation. It looks a little bit different.

$$K(r) = \sigma^2(1 + \sqrt{5}r + \frac{5}{3}r^2)\exp(-\sqrt{5}r)$$

It looks like, the positive parameter l is set to 1. As additional hyper parameter the variance, σ^2, was introduced.

1.4 Hyper-Parameters

The *GPy* library has for every kernel a variance parameter. This parameter is a hyper parameter to adjust improve the prediction result. If I set the parameter to a lower value the values for the learning are in a smaller gap. If I increase σ^2 the gab increases. So if the variance of my data is high, an increased value for the parameter variance makes sense.

1.5 Implementation

1.6 Result

1.7 Conclusion

Bibliography

Guo-Xun Yuan, Chia-Hua Ho and Chih-Jen Lin (2021). "Recent Advances of Large-Scale Linear Classification". In: *Proceedings of the IEEE* 100, pp. 2584–2603.

KOWALCZYK, Alexandre (2014). *Linear Kernel: Why is it recommended for text classification*. URL: `https://www.svm-tutorial.com/2014/10/svm-linear-kernel-good-text-classification/`.

Powell, Alex (2021). *Multivariate normal distribution*. URL: `https://towardsdatascience.com/intro-to-gaussian-process-regression-14f7c647d74d`.

Rueckert, Elmar (2022). *An Introduction to Probabilistic Machine Learning*. Elmar Rueckert.

Sreenivasa, Sushanth (2020). *Radial Basis Function (RBF) Kernel: The Go-To Kernel*. URL: `https://www.svm-tutorial.com/2014/10/svm-linear-kernel-good-text-classification/`.

Williams, C. E. Rasmussen C. K. I. (2006). *Gaussian Processes for Machine Learning*. MIT Press, p. 85.

APPENDIX