Data Science Lab 2024: Exercise sheet 5 (b)

April 29, 2024

1 Bayesian linear regression

In the lecture we looked into a Bayesian linear regression model. For the linear model described in the lecture show how the Bayesian regression fit changes with the training set sample size (size of 50 and 1000). In both cases draw samples from the posterior distribution. Do a PIP install to install PYMC3 for this exercise.

2 Stellar activity I

Most exoplanet host stars are magnetic in nature and their magnetic activity is a big source of noise in exoplanet detection tecniques.

(a) For the planet host HD166724 load the stellar activity (s-index) data from the attached file and plot the activity as a function of time. Use NUMPY to load the data and MATPLOTLIB or SEABORN for plotting.

(b) Apply an appropriate kernel* (assuming a zero mean Gaussian). The data has to be centered at zero to apply this assumption. Explain your kernel of choice.

(c) Use Gaussian processes to draw a few samples from the posterior distribution and plot the mean predictions.

(d) Explore how the hyperparameters of the kernal influences the confidence level of the mean predictions.

*we discussed squared exponential kernel in the lecture. You can use any other kernel.

3 Stellar activity II

Same as exercise 2 (a) to (d) but on a different dataset: Sun.csv. What does the mean of your posterior distribution tells you about the temporal evolution of the data? Is there any resemblance to the 11 year sunspot cycle? Explain your answer in a few sentences.

NOTE: The time axis is given in Julian Dates in the attached files. The HD166724.csv file contains multiple different columns. Only the Julian date columns and the s-index

columns are of interest in this exercise. You can use tjd or bjd as you see appropriate. The Sun.csv dataset only has two columns. The julian date and the s-index. If you are more comfortable with calendar dates use ASTROPY to convert your time axis into appropriate units.

In the lecture we discussed the plain PYTHON implementation of Gaussian processes. You can also use existing libraries such as SCIKIT-LEARN or PYMC3 or any other python package of your choice.