

FYSS5120 Efficient Numerical Programming - Demo 5

Drop solutions before the demo session to the Nextcloud box ([link](#))
Please indicate clearly your name in the file name,
so that I can tell who solved what.

1. The file `measured_data.txt` contains three x, y values,

x	y
0.1	2.1
1.0	3.2
9.2	1.8

Write a code that does gaussian process regression to this data set, plot the resulting mean function and 95% confidence limits. You can modify the code `sklearn_GPR.py` in directory `numerics` ([link](#)).

2. Why does the mean function goto to zero between the second and third data point?
3. After all this work your boss walks in and says “the result should be close to the curve $y = \sin(x) + 2$ ”. That changes the prior to a new value. Modify the code and repeat the calculation. Hint: rather than modifying the underlying `sklearn.GaussianProcessRegressor`, write a function

```
def mod_prior(x):  
    return np.sin(x) + 2
```

and subtract the oscillation `mod_prior(x)` from the three data points. Proceed with GPR as usual. After all is done, add back oscillation to all results and plot them as earlier. Only thing you have to pay attention to is what y 's to modify with what input x 's to `mod_prior(x)`. There are array shapes `(51,)` and `(51,1)` around...

4. Voluntary: Try out the parameter `alpha` in `sklearn.GaussianProcessRegressor`. Setting it to a small value adds noise to the fitting procedure.
5. Voluntary: What happens if the second point x is close to the first one? Why do you get messy results?