NO AI USED

Date
Page

Gaussian Process Regression

Let us first consider a function as given $f(x)$ for any x. In this case, for an example I am considering a toy dataset as given in the question.
So let $f(x) = x+2$ at $x=1$, $x=2$, $x=3$
If I consider the extrapolation at $x'=4$. I am not considering any noise in this case.
So, we can consider $y = f(x) = [3, 4, 5]$
From the text, I understood that,
For prediction, let us apply a kernel.
For GPR, kernel is defined as

$$k(x, x') = \sigma_f^2 e^{\left(-\frac{(x-x')^2}{2l^2}\right)}$$

and
$$K = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & k(x_1, x_3) \\ k(x_2, x_1) & k(x_2, x_2) & k(x_2, x_3) \\ k(x_3, x_1) & k(x_3, x_2) & k(x_3, x_3) \end{bmatrix}$$

$$K_* = [k(x', x_1), k(x', x_2), k(x', x_3)]$$

$$K_{**} = k(x', x')$$

So, let the predicted function be $F(x)$.
So, we define
the best fit of y as $\bar{y}_*$ (which is mean of y)
is defined as

$$\bar{y}_* = K_* K^{-1} \bar{y} \quad \text{and} \quad var(y_*) = K_{**} - K_* K^{-1} K_*^T$$

Date
Page

Here, I have two hyper parameters.
they are σ_F & l.

σ_F is most significant for getting the "spread".
here σ_f corresponds to the maximum allowed covariance

In this case, since the gaussian noise is not present
I will consider σ_F to be small, say 0.1.

l is the length (or distance) between the known point and the ~~pre~~ x' value to be used for prediction. Since, the distance is $x-x'$ is not so small compared to the spread of x, I will consider it to be relatively equal to the size of the spread of x. say $l = 10$.

In case we have some more spread in y values along with some noise, we may use the formula ~~from the~~ for the kernel as

$$k(x, x') = \sigma_F^2 \exp\left[\frac{-(x-x')^2}{2l^2}\right] + \sigma_n^2 \delta(x, x')$$

Where δ = Kronecker δ fn.

The implementation is in the colab notebook as well as the second part ~~to~~ (and the bonus part) are present in it as well as this text (previous & this page).