2020 Machine Learning Homework 5

School: National Ciao Tung University
Name: Shao-Wei (Willy) Chiu
Student ID: 0856090

Gaussian Process

- Training data:
  - $\begin{bmatrix} x_1 & y_1 \\ . & . \\ . & . \\ . & . \\ x_n & y_n \end{bmatrix} = [X \quad Y]$
- Kernel function:
  - $k(x_n, x_m) = \sigma^2 (1 + \frac{\|x_n - x_m\|^2}{2\alpha\ell^2})^{-\alpha}$

There is a function f could transfer each x_i into coressponding y_i (i.e., $f(x_i) = y_i$).
Assume that $y_i = f(x_i) + \epsilon, where\ \epsilon \sim N(0, \beta)$ and $f \sim N(0, K_n)$.
(i.e., $Y \sim N(f, \beta)$)
On estimate the x_* point, we have formula $\begin{bmatrix} Y \\ y_* \end{bmatrix} \sim N(\begin{bmatrix} Y \\ y_* \end{bmatrix} |0, K_{n+1})$

After the derivation of probability, we get the

- $\mu(x_*) = k(x, x_*)^T (K_n + \beta I)^{-1} Y$
- $cov(x_*) = k(x_*, x_*) - k(x, x_*)^T (K_n + \beta I)^{-1} k(x, x_*)$

> This form is almost the same as the formula mentioned by Prof. Chiu in the class.
> The tiny difference is that it take the β out of the matrix K, but the course silde takes it into the matrix.

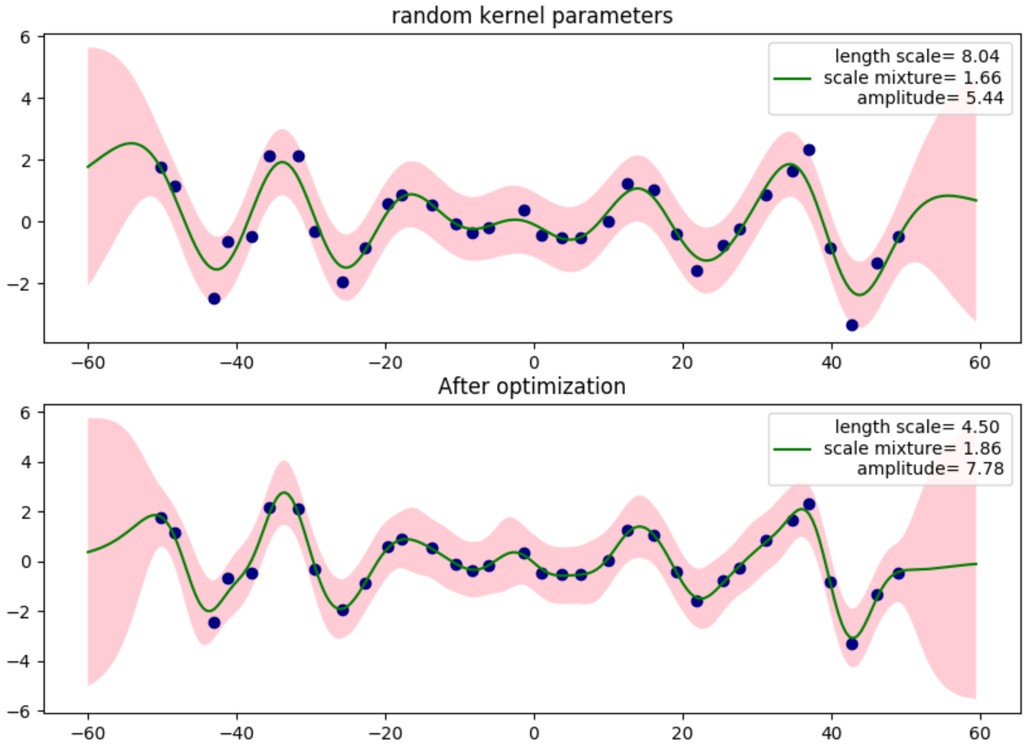
Thus, we could apply the x_* to describe our model.

We notice that the kernel method is decided by some kernel parameters (e.g., σ, α, ℓ), so we need to find the parameters which could have the maximum likelihood.

In my practice, I choose the random value of all parameter between 0 and 10, and call the `scipy.optimize.minimize` to optimize it.
The relative formula is shown below,

$$\begin{aligned} argmax(ln\ p(y \mid \theta)) &= -\tfrac{1}{2} ln\ |C_\theta| - \tfrac{1}{2} y^T C_\theta^{-1} - \tfrac{N}{2} ln\ (2\pi) \\ &\propto -ln\ |C_\theta| - y^T C_\theta^{-1} y \\ &= argmin(\ ln\ |C_\theta| + y^T C_\theta^{-1} y\ ) \end{aligned}$$

result:



Referrence

- https://www.csie.ntu.edu.tw/~cjlin/mlgroup/tutorials/gpr.pdf