1. code with detailed explanations

Part 1

The rational quadratic kernel is

$$k(x_a, x_b) = \sigma^2 \left(1 + \frac{\|x_a - x_b\|^2}{2\alpha\ell^2}\right)^{-\alpha}$$

And I use the kernel to calculate mean and varience

$$\mu(\mathbf{x}^*) = k(\mathbf{x}, \mathbf{x}^*)^\top \mathbf{C}^{-1}\mathbf{y}$$
$$\sigma^2(\mathbf{x}^*) = k^* - k(\mathbf{x}, \mathbf{x}^*)^\top \mathbf{C}^{-1} k(\mathbf{x}, \mathbf{x}^*)$$
$$k^* = k(\mathbf{x}^*, \mathbf{x}^*) + \beta^{-1}$$

After I get the mean and varience, I can visualize the result
The black line is the mean and the confidence interval is mean add varience / minus varience

Part 2

To optimize the kernel parameter, I need to minimize negative marginal log-likelihood function
Marginal log-likelihood function is

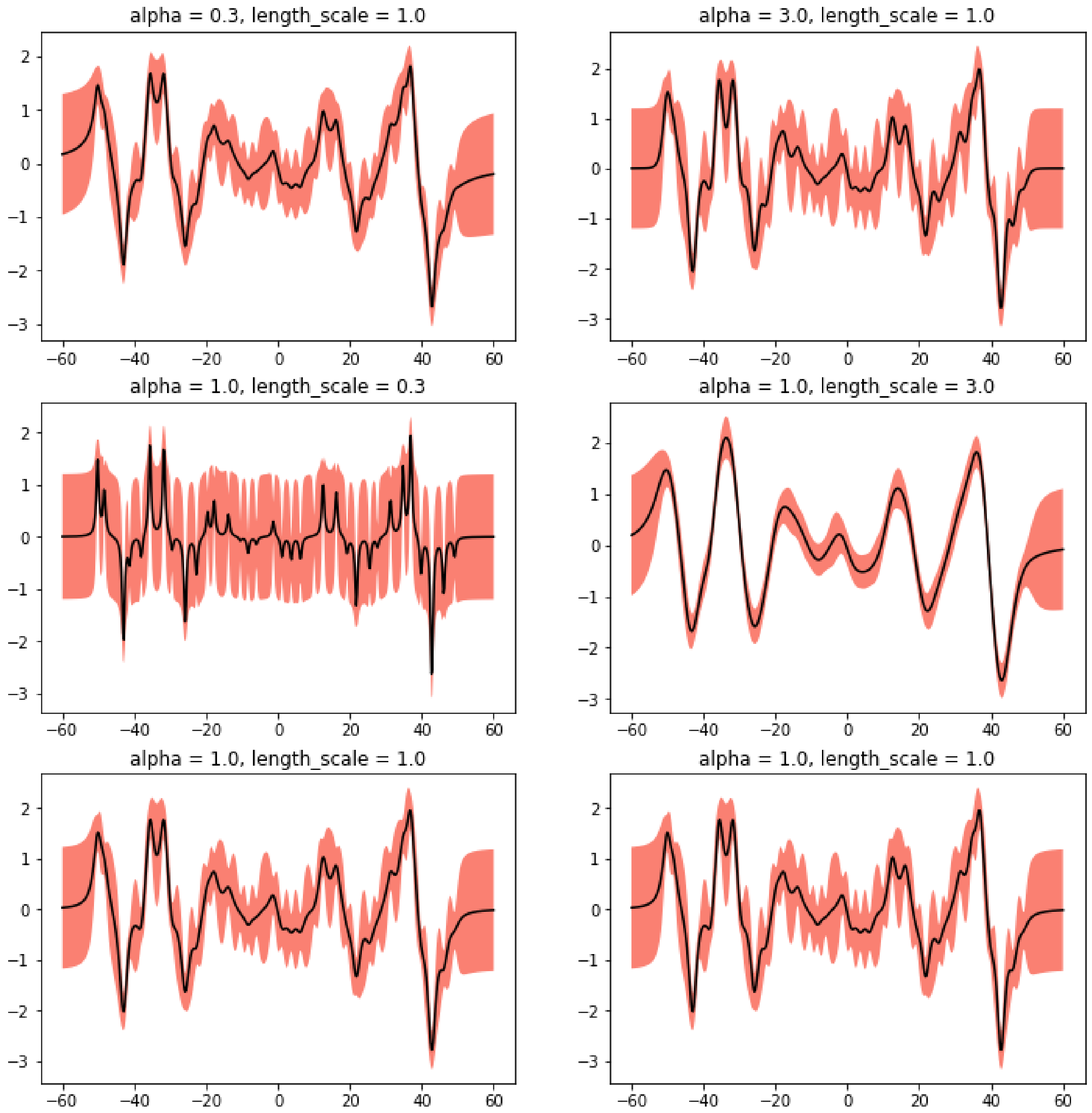
$$p(\mathbf{y}|\theta) = \mathcal{N}(\mathbf{y}|\mathbf{0}, \mathbf{C}_\theta)$$
$$\ln p(\mathbf{y}|\theta) = -\frac{1}{2}\ln\ |\mathbf{C}_\theta| - \frac{1}{2}\mathbf{y}^\top \mathbf{C}_\theta^{-1}\mathbf{y} - \frac{N}{2}\ln\ (2\pi)$$

I use the scipy.optimize.minimize to optimize the parameters. Then, when I get the optimal parameter, I can calculate the new kernel, mean, varience.

2. Experiments settings and results

Part 1

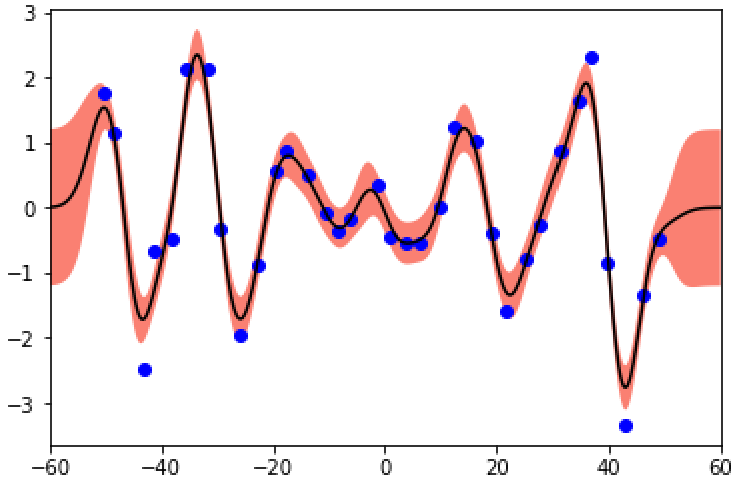
The image shows the result of applying Gaussian Process Regression to predict the distribution w/o optimizing the kernel parameters
The title of images shows the hyperparameter setting



Part 2

The image shows the result of applying Gaussian Process Regression to predict the distribution w/ optimizing the kernel parameters

hyperparameter initial guess
alpha=1,length_scale=1



3. Observations and discussion

Higher length_scale lead to smoother functions. Lower length_scale make functions more wiggly with wide confidence interval.
Alpha controls the vertical variation of functions drawn from the GP. This can be seen by the wide confidence intervals outside the training data region