Machine Learning Homework 5 - 1

Gaussian Process

Due Date 23:55 23th Nov.

I. Gaussian Process

In this section, you are going to implement the Gaussian Process and visualize the result.

- Training data
  - **input.data** is a 34x2 matrix. Every row corresponds to a 2D data point (X_i,Y_i).
  - $Y_i = f(X_i) + \epsilon_i$ is a noisy observation, where $\epsilon_i \sim N(\cdot|0, \beta^{-1})$. You can use $\beta = 5$ in this implementation.
- What you are going to do
  - Part1: Apply Gaussian Process Regression to predict the distribution of *f* and visualize the result. Please use a rational quadratic kernel to compute similarities between different points.
    Details of the visualization:
    - Show all training data points.
    - Draw a line to represent the mean of *f* in range [-60,60].
    - Mark the 95% confidence interval of *f*.

    (You can use matplotlib.pyplot to visualize the result, e.g. use matplotlib.pyplot.fill_between to mark the 95% confidence interval, or you can use any other package you like.)
  - Part2: Optimize the kernel parameters by minimizing negative marginal log-likelihood, and visualize the result again. (You can use scipy.optimize.minimize to optimize the parameters.)

II. Report

- Submit a report in pdf format. The report should be written in English.
- Report format:
  - 1. code with detailed explanations (20%)
    - For example, show the formula of rational quadratic kernel and the process you optimize the kernel parameters
    - Note that if you don't explain your code, you cannot get any points in section 2 and 3 either.
      - Part1 (10%)
      - Part2 (10%)
  - 2. experiments settings and results (20%)
    - Show the figures and the hyperparameters we asked you to show
    - Note that if you don't explain your code in the above section, you cannot get any points in this section either.
      - Part1 (10%)
      - Part2 (10%)
  - 3. observations and discussion (10%)
    - Anything you want to discuss, such as comparing the performance when using different hyperparameters.

III. Turn in

1. Report (.pdf)
2. Source code

You should zip source code and report in one file and name it like ML_HW5-1_yourstudentID_name.zip, e.g. ML_HW5-1_0856XXX_王小明.zip.

P.S. If the zip file name has format error or the report is not in pdf format, there will be a penalty (-10). Please submit your homework before the deadline, late submission is not allowed.

♦ Packages allowed in this assignment:

You are only allowed to use numpy, scipy.optimize, scipy.spatial.distance, and package for visualizing results. Official introductions can be found online.
Important: scikit-learn is not allowed.