Introduction to Machine Learning
Lab 4: Kernel Regression

Hongteng Xu

March 27, 2022

1 Motivation

- Implement some typical kernel functions, and check their feasibility in different models. You may find that some kernels work better for some models while perform poorly for the other models:)
- Implement classic kernel methods like Nadaraya-Watson estimator and kernel ridge regression (KRR). Observe their differences and connections.
- Take KRR as a good example. Try to sharpen your hyperparameter fine-tuning feelings.

2 Tasks

Please read Lecture 5 carefully before doing this lab work.

1. Implement the following four kinds of kernel functions and use them to achieve the Nadaraya-Watson estimator shown in the lecture.

$$\begin{aligned}&\text{RBF: } K_h(\boldsymbol{x},\boldsymbol{x}')=\exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{x}'\|_2^2}{h}\right), \quad \text{Gate: } K_h(\boldsymbol{x},\boldsymbol{x}')=\begin{cases}\frac{1}{h}, & \|\boldsymbol{x}-\boldsymbol{x}'\|_1\leq h\\ 0 & \text{Otherwise.}\end{cases}\\ &\text{Triangle: } K_h(\boldsymbol{x},\boldsymbol{x}')=\begin{cases}\frac{2}{h}(1-\frac{\|\boldsymbol{x}-\boldsymbol{x}'\|_1}{h}) & \|\boldsymbol{x}-\boldsymbol{x}'\|_1\leq h\\ 0 & \text{Otherwise}\end{cases}, \quad \text{Linear: } K(\boldsymbol{x},\boldsymbol{x}')=\boldsymbol{x}^T\boldsymbol{x}'\end{aligned} \tag{1}$$

   where h is the hyperparameter. The NW estimator: $y_{new}=\sum_{n=1}^{N}\frac{K(\boldsymbol{x}_{new},\boldsymbol{x}_n)}{\sum_{i=1}^{N}K(\boldsymbol{x}_{new},\boldsymbol{x}_i)}y_n$.

2. Implement the closed-form solution of kernel ridge regression:

$$\min_{\boldsymbol{a}}\|\boldsymbol{y}-\boldsymbol{K}\boldsymbol{a}\|_2^2+\tau\boldsymbol{a}^T\boldsymbol{K}\boldsymbol{a}. \tag{2}$$

   (Hint: Use as few computations as possible.)

3. **Struggle with KRR's SGD:** Implement a stochastic gradient descent (SGD) algorithm to solve (2). (Hint: Do your best to make it work, and think about whether it actually "works" or not.)