CS 391L: Machine Learning **Fall 2023**

Homework 6 - Programming

Lecture: Prof. Qiang Liu

1. **Kernel Regression**

   Given a training data $\{\boldsymbol{x}_i, y_i\}_{i=1}^n$, kernel regression approximates the unknown nonlinear relation between $\boldsymbol{x}$ and y with a function of form

   $$y \approx f(\boldsymbol{x};\ \boldsymbol{w}) = \sum_{i=1}^n w_i k(\boldsymbol{x},\ \boldsymbol{x}_i),$$

   where $k(\boldsymbol{x}, \boldsymbol{x}')$ is a positive definite kernel specified by the users, and $\{\boldsymbol{w}_i\}$ is a set of weights, estimated by minimizing a regularized mean square error:

   $$\min_{\boldsymbol{w}} \left\{ \sum_{i=1}^n (y_i - f(\boldsymbol{x};\ \boldsymbol{w}))^2 + \beta \boldsymbol{w}^\top \boldsymbol{K} \boldsymbol{w} \right\},$$

   where $\boldsymbol{w}$ is the column vector formed by $\boldsymbol{w} = [w_i]_{i=1}^n$ and $\boldsymbol{K}$ the $n \times n$ matrix by $\boldsymbol{K} = [k(\boldsymbol{x}_i, \boldsymbol{x}_j)]_{ij=1}^n$, and β is a positive regularization parameter. We use a simple Gaussian radius basis function (RBF) kernel,

   $$k(\boldsymbol{x}, \boldsymbol{x}') = \exp\left(-\frac{||\boldsymbol{x} - \boldsymbol{x}'||^2}{2h^2}\right),$$

   where h is a bandwith parameter. A common way to set h in practice is the so called "median trick", which set h to be the median of the pairwise distance on the training data, that is,

   $$\hat{h}_{\text{med}} = \text{median}(\{||\boldsymbol{x}_i - \boldsymbol{x}_j|| \colon i \neq j,\ i, j = 1, \ldots, n\}).$$

   (1) [**10 points**] Complete the code of kernel regression following the instruction of the attached Python notebook. Specifically, you need to complete all the code necessary for function
   `kernel_regression_fit_and_predict` to run.

   (2) [**10 points**] Run the algorithm with $\beta = 1$ and $h \in \{0.1\hat{h}_{\text{med}},\ \ \hat{h}_{\text{med}},\ \ 10\hat{h}_{\text{med}}\}$. Show the curve learned with different h in the notebook and comment on how h influences the smoothness of the curve.

   (3) [**10 points**] Use 5-fold cross validation to find the optimal combination of h and β within $h \in \{0.1\hat{h}_{\text{med}},\ \ \hat{h}_{\text{med}},\ \ 10\hat{h}_{\text{med}}\}$ and $\beta \in \{0.1, 1\}$.