

## CMSE 820 HW8

This HW is due on Nov 10th at 11:59pm.

### Question 1:

- (1) Write a function to solve the following RKHS regression problem.

$$\hat{f} = \operatorname{argmin}_{f \in \mathcal{H}_k} \sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \|f\|_k^2,$$

where  $\|\cdot\|_K$  is the norm defined in the RKHS with Gaussian kernel  $K(x, y) = \exp(-\|x - y\|^2/0.25)$ , Laplacian kernel  $K(x, y) = \exp(-\|x - y\|)$ , or Polynomial kernel  $K(x, y) = (\langle x, y \rangle + 1)^2$ .

- (2) Download data from D2L, and test your function with Gaussian and Laplacian kernels. You need to identify the best  $\lambda$  using 6-fold cross-validation. Plot  $x$  vs  $f(x)$  on top of the original data points.
- 3 (optional) Try smoothing spline with cross-validation to find your best  $\lambda$  and report your result.

**Question 2:** Let's play with the MNIST data set. Consider again a binary classification problem between pairs of numbers, e.g, 1 and 4, 3 and 5, etc. Take the labels to be  $Y = \{+1, -1\}$ , +1 for one digit, -1 for the other digit. We can use the kernel regression algorithm to define a classifier. Indeed, a class label can be defined as:

$$y = \operatorname{sgn} f(x) = \operatorname{sgn} \left( \sum_{i=1}^n \alpha_i k(x_i, x) \right).$$

In this case, the magnitude  $|f(x)|$  can be interpreted as a measure of confidence in the prediction - the larger the value, the more confident the prediction. Train such a regression with the quadratic polynomial  $k(x, x') = (\langle x, x' \rangle + c)^2$  using RKHS ridge regression. Report your results in terms of accuracy in the testing set.

- 1: Do this for all pairs of 1, 2, 3, 4.
- 2: You will have to do a 60% training, 40% testing random split, because otherwise the training kernel  $K$  can get quite large in memory.
- 3: (optional) Try different kernels and compare their performance.

**Question 3:** Prove the Semiparametric Representer Theorem: Let  $c : \mathcal{X} \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R} \cup \{\infty\}$  be a cost function,  $\Omega : [0, \infty) \rightarrow \mathbb{R}$  a strictly monotonic increasing function, and  $\mathcal{H}_k$  a RKHS. Additionally, let  $\{\psi_j : \mathcal{X} \rightarrow \mathbb{R}\}_{j=1}^m$  be a collection of  $m$  real valued functions with

the property that the  $n \times m$  matrix  $(\psi_j(x_i))_{ij}$  has rank  $m$ . Finally, let  $\mathcal{F}$  be the functional class:

$$\mathcal{F} = \{\tilde{f} = f + h : f \in \mathcal{H}_k, \quad h \in \text{Span}\{\psi_j\}_{j=1}^m\}.$$

Then each minimizer:

$$\tilde{f}^* = \operatorname{arginf}_{\tilde{f} \in \mathcal{F}} \sum_{i=1}^n \frac{1}{n} c[x_i, y_i \tilde{f}(x_i)] + \Omega(\|\tilde{f}\|_{\mathcal{H}_k}),$$

admits a representation of the form:

$$\tilde{f}^*(x) = \sum_{i=1}^n \alpha_i k(x_i, x) + \sum_{j=1}^m \beta_j \psi_j(x), \quad \alpha_i, \beta_j \in \mathbb{R}.$$

**Question 4:** Given a PSD kernel  $k$ , the normalized kernel defined on  $\mathcal{X}$  is

$$\tilde{k} = \frac{k(x, x')}{\sqrt{k(x, x)k(x', x')}}.$$

Prove  $\tilde{k}$  is a p.s.d. kernel.

**Question 5:** Let  $k(\cdot, \cdot)$  be a PSD kernel satisfying the conditions in Mercer's theorem. Define

$$\mathcal{H} = \{f = \sum_{j=1}^N w_{[j]} \sqrt{\lambda_j} \psi_j : \{w_{[j]}\} \in \ell_2\}.$$

1. Solve the following problem:

$$f = \arg \min_{f \in \mathcal{H}} \sum_{j=1}^n (y_i - f(x_i))^2 + \lambda \|w\|^2,$$

where  $w = (w_{[1]}, w_{[2]}, \dots, w_{[N]})^T$  and  $\|w\|^2 = w^T w$ .

2. Show that this formulation is equivalent to RKHS ridge regression.