9.520 Problem set 1

Pr. 1.1 Reproducing Kernel Hilbert Spaces

1. An RKHS can be defined via a kernel K and has the following reproducing property:
$$\mathcal{F}_t[f] = \langle K(t, \cdot), f(\cdot) \rangle_K = f(t).$$
Given two functions
$$\begin{aligned} f(x) &= \sum_{i=1}^{\ell} a_i K(x, x_i) \\ g(x) &= \sum_{i=1}^{\ell} b_i K(x, x_i) \end{aligned}$$
what is $\langle f, g \rangle_K$, $||f||_K^2$, and $||g||_K^2$?
2. Given Mercer's theorem
$$K(s, t) = \sum_{n=1}^{\infty} \lambda_n \phi_n(s) \phi_n(t),$$
where $\lambda_n \geq 0$ and the following definition of a RKHS norm for functions $f(x) = \sum_n c_n \phi_n(x)$
$$||f||_k^2 = \sum_{n=1}^{\infty} \frac{c_n^2}{\lambda_n}$$
where $\phi_n(x)$ are the eigenfunctions of the integral operator defined by the kernel.
Prove that the reproducing property holds.
Prove that one gets the same form for the quantities $\langle f, g \rangle_K$, $||f||_K^2$, and $||g||_K^2$ as obtained in part 1 where
$$\begin{aligned} f(\cdot) &= \sum_{n=1}^{\infty} c_n \phi_n(\cdot) \\ g(\cdot) &= \sum_{n=1}^{\infty} d_n \phi_n(\cdot) \end{aligned}$$
and compute the relation between the a_i and the c_n.

Pr. 1.2 The discrete counterpart in $\mathbb{R}^n$ of the eigenfunction equation
$$\int_0^1 K(x, y) v(y) d\mu(y) = \sigma v(x),$$
where $K(x, y)$ is a positive definite function and μ a suitable measure, is
$$KP\mathbf{v} = \sigma \mathbf{v} \tag{1}$$
with K a positive definite matrix and P a diagonal positive definite matrix (basically a weighting on each data point).

1. Show that there exists a solution to equation (1) of n linearly independent vectors $\mathbf{v}_i$ and corresponding strictly positive values σ_i. (*Hint:* Consider the matrix $P^{1/2}KP^{1/2}$...)
2. Let $\mathbf{f} \in \mathbb{R}^n$ be represented as
$$\mathbf{f} = \sum_{i=1}^{n} a_i \mathbf{v}_i$$
with
$$a_i = \mathbf{f}^\top P \mathbf{v}_i.$$
Prove that the norm of $\mathbf{f}$
$$||\mathbf{f}||_k^2 = \sum_{i=1}^{n} \frac{a_i^2}{\sigma_i}$$
is independent of P. What can you conclude about the dependence on the measure P of the norm of $\mathbf{f}$ in the RKHS?

Pr 1.3 Given a training set of points not necessarily linearly separable consider the SVM obtained by minimizing
$$\frac{1}{2}\lambda\|\mathbf{w}\|^2 + \frac{1}{2}\sum_{i=1}^{\ell} \xi_i^2,$$
$\lambda > 0$, with respect to $\mathbf{w}$, b, and $\boldsymbol{\xi}$ subject to the constraints
$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, \quad i = 1, ..., \ell.$$

1. Derive the dual formulation and show that the associated QP is equivalent to the QP discussed in class in the linearly separable case but with the Kernel matrix with entries defined as
$$K_{ij} = \frac{1}{\lambda} y_i y_j \mathbf{x}_i \cdot \mathbf{x}_i + \delta_{ij}.$$
2. Set $b = 0$ and show that if the ℓ inequalities become equalities the learning scheme is Regularized Least-Squares Classification. Derive the associated linear system of equations and compare the matrix which has to be inverted with the Kernel matrix above.

Pr 1.4 For the SVM classification problem we have the following optimality conditions

$$\sum_{j=1}^{\ell} c_j K(\mathbf{x_i}, \mathbf{x_j}) - \sum_{j=1}^{\ell} y_i \alpha_j K(\mathbf{x_i}, \mathbf{x_j}) = 0 \qquad i = 1, \ldots, \ell$$
$$\sum_{i=1}^{\ell} \alpha_i y_i = 0$$
$$C - \alpha_i - \zeta_i = 0 \qquad i = 1, \ldots, \ell$$
$$y_i(\sum_{j=1}^{\ell} y_j \alpha_j K(\mathbf{x_i}, \mathbf{x_j}) + b) - 1 + \xi_i \geq 0 \qquad i = 1, \ldots, \ell$$
$$\alpha_i[y_i(\sum_{j=1}^{\ell} y_j \alpha_j K(\mathbf{x_i}, \mathbf{x_j}) + b) - 1 + \xi_i] = 0 \qquad i = 1, \ldots, \ell$$
$$\zeta_i \xi_i = 0 \qquad i = 1, \ldots, \ell$$
$$\xi_i, \alpha_i, \zeta_i \geq 0 \qquad i = 1, \ldots, \ell.$$

Derive the following "reduced" optimality conditions:

$$\begin{aligned} \alpha_i = 0 \quad &\text{“}\Longleftrightarrow\text{”} \quad y_i f(x_i) \geq 1 \\ 0 < \alpha_i < C \quad &\text{“}\Longleftrightarrow\text{”} \quad y_i f(x_i) = 1 \\ \alpha_i = C \quad &\text{“}\Longleftrightarrow\text{”} \quad y_i f(x_i) \leq 1 \end{aligned}$$

and explain why we put quotes around $\Longleftrightarrow$.

Pr 1.5 We are given the SVM for regression with the ϵ-insensitive loss function without a b term

$$\min_{f \in \mathcal{H}} \frac{1}{\ell} \sum_{i=1}^{\ell} (|f(\mathbf{x}) - y| - \epsilon)_+ + \lambda \|f\|_K^2.$$

1. Write the primal formulation with slack variables (*Hint:* You need two slack variables for the error at each point instead of one as was the case in classification).
2. Derive the Lagrangian (*Hint:* You need a multiplier for each slack variable).
3. Derive the dual formulation.

Pr 1.6 Consider any function of one variable that is continuous, symmetric and periodic with positive Fourier coefficients. Then $K(x)$ can be expanded in a uniformly convergent Fourier series (all normalization factors set to 1):

$$\begin{aligned} K(x) &= \sum_{n=0}^{\infty} \lambda_n cos(nx) \\ K(x-y) &= 1 + \sum_{n=1}^{\infty} \lambda_n \sin(nx)\sin(ny) + \sum_{n=1}^{\infty} \lambda_n \cos(nx)\cos(ny), \end{aligned}$$

the eigenvalues are set to $\lambda_n = 2^{-n}$ and the eigenfunctions of K are

$$(1, \sin(x), \cos(x), \sin(2x), \cos(3x), ..., \sin(px), \cos(px), ...).$$

The RKHS norm of this function is

$$\|f\|_K^2 \equiv \sum_{n=0}^{\infty} \frac{\langle f, \cos(nx)\rangle^2 + \langle f, \sin(nx)\rangle^2}{\lambda_n} < \infty.$$

1. Show that if we write the functions as

   $$f(x) = \sum_{n=1}^{\infty} c_n \sin(nx) + d_n \cos(nx)$$

   then the coefficients c_n and d_n are bounded as follows

   $$c_n^2 + d_n^2 < 2^{-n}.$$

2. Prove that the space of functions spanned by this RKHS, functions spanned by

   $$f(x) = \sum_{n=1}^{\infty} c_n \sin(nx) + d_n \cos(nx)$$

   with $||f||_K < \infty$ is compact. (*Hint:* Look at the proof in Mathcamp 1 of the compactness of the Hilbert cube and use the fact that $c_n^2 + d_n^2 < 2^{-n}$.).

Pr. 1.7 An algorithm is (β, δ) Cross-Validation (CV) stable if for almost all $S \in \mathcal{Z}^\ell$ (except for a set of measure δ), the following holds:

$$\forall i, u \in \mathcal{Z},\ |c(f_S, u) - c(f_{S^{i,u}}, u)| \leq \beta.$$

Assuming that our algorithm $\mathcal{A}$ has CV-stability β and assuming that the loss function $c(f, z)$ (equal to $V(f(\mathbf{x}), y) \geq 0$) non-negative and bounded above my M, show that

$$|\mathbb{E}_S D[f_S]| \leq \beta + M\delta.$$

Here $D[f_S]$ is the defect defined in lectures (empirical error - expected error). Note that bounding the expectation of the defect was one of two things we had to show for application of McDiarmid's inequality.