GI13 – Problem set (Due Thursday 12am, March 24)

Note: Presentation, clarity, and synthesis of exposition will be taken into account in the assessment of these exercises.

1. **[20 pts]** (*representer theorem*)

   Consider the ridge regression optimisation problem

   $$\min_{w \in \mathbb{R}^N} \left\{ \sum_{i=1}^{m} (w^\top \phi(x_i) - y_i)^2 + \gamma \|w\|^2 \right\}, \tag{1}$$

   where $\|\cdot\|$ denotes the L_2 norm, $\gamma > 0$ is a fixed real number, $\phi : \mathbb{R}^d \to \mathbb{R}^N$ is a feature map, $x_1, \dots, x_m \in \mathbb{R}^d$ are given inputs and $y_1, \dots, y_m \in \mathbb{R}$ given outputs.

   Using the Representer Theorem, derive a problem equivalent (dual) to (1), which involves only the Gram matrix (and does not involve the inputs x_i). Assume that the Gram matrix is invertible. Then, derive the optimal solution of the dual problem. Let $\hat{w}$ be the solution of problem (1). Write an expression for the function

   $$x \mapsto \hat{w}^\top \phi(x)$$

   in terms of the kernel K defined as $K(x,t) = \phi(x)^\top \phi(t)$.

2. **[15 pts]** (*kernels*)

   Consider the Gaussian kernel $K(x,t) = \exp(-\beta(x-t)^2)$, for $x, t \in \mathbb{R}$. Find a feature map representation for this kernel (*hint:* use the result for the infinite polynomial kernel).

3. **[15 pts]** (*ℓ_1-norm interpolation*)

   Consider the minimal ℓ_1-norm interpolation problem $\min\{\|w\|_1 : w \in \mathbb{R}^d,\ w^\top x_i = y_i,\ i = 1, \dots, m\}$, where $w \in \mathbb{R}^d$ is a vector of parameters we minimize over, $\|w\| = \sum_{i=1}^n |w_i|$, and $(x_1, y_1), \dots, (x_m, y_m) \in \mathbb{R}^d \times \mathbb{R}$ are given datapoints which can be interpolated.

   (a) Show that this problem is linear programming problem.

   (b) Consider the case that $d = 2$ and $m = 1$. Give an instance (that is, provide the value (x_1, y_1) of the interpolating datapoint) for which the optimization problem has a unique solution.

4. **[8 pts]** (*trace norm regularization*) Let $W \in \mathbb{R}^{d \times n}$ be a $d \times n$ matrix. Write a singular value decomposition of W and then prove that

   $$\sum_{i=1}^{d} \sum_{j=1}^{n} W_{ij}^2 = \mathrm{tr}(W^\top W) = \mathrm{tr}(W W^\top) = \sum_{i=1}^{r} \sigma_i^2$$

   where W_{ij} denotes the (i,j) element of W, $\sigma_1, \dots, \sigma_r$ are the singular values of W and r is the rank of W.

5. **[12 pts]** (*convex functions*)

   Let $f : \mathbb{R}^d \to \mathbb{R}$ be a *convex* function such that

   $$f(w) \geq 0 \qquad \text{for every } w \in \mathbb{R}^d.$$

   Show that the function $g : \mathbb{R}^d \to \mathbb{R}$ defined as

   $$g(w) = (f(w))^2 \qquad \text{for every } w \in \mathbb{R}^d$$

   is also convex. As a special case deduce that, for every norm $\|\cdot\|$, the function

   $$w \mapsto \|w\|^2$$

   is convex.

6. **[10 pts]** (*convexity*)

   Let S^n_{++} be the open cone of positive definite matrices. Prove that the function $f : \mathbb{R}^d \times S^d_{++}$ defined as $f(z, A) \mapsto z^\top A^{-1} z$ is convex.

7. **[10 pts]** (*linear programming*)

   Solve the following linear programming problem

$$\begin{aligned} \min_{x,y \in \mathbb{R}} \quad & 2y - x \\ \text{subject to} \quad & 2x - y \leq -1 \\ & -3x - y \leq 0\,. \end{aligned}$$

   Hint: Solve the inequalities for y, then replace y in the objective and minimise over x; consider cases if it helps.

8. **[10 pts]** (*convex optimization*) Consider the optimisation problem

$$\min_{w \in \mathbb{R}^d} \sum_{i=1}^{m} \max\{w_i^2 + 2w_i - 5, 0\} + \|w\|^2\,,$$

   where w_i denotes the i-th component of w. Is this a convex program? Assume that a minimiser exists. Is the minimiser unique or not? Rewrite the problem in the standard form of a well known type (LP, QP, QCQP, SDP).

9. **BONUS 1 [5 pts]** (*kernels*)

   Let $x, t \in [-1, 1]$ and define the kernel

$$K(x, t) = \frac{1}{1 - xt}.$$

   (a) **[3 pts]** Show that K is a valid kernel.

   (b) **[2 pts]** Given any distinct inputs $x_1, \ldots, x_m \in [-1, 1]$ show that the kernel matrix $\mathbf{K} = (K(x_i, x_j) : i, j = 1, \ldots, m)$ is invertible.