Advanced Topics in Machine Learning

Assignment 2

1. This question pertains to regularisation methods with kernels. Consider the following optimisation problem (ridge regression)

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

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, $x_1, \ldots, x_m \in \mathbb{R}^d$ are given inputs and $y_1, \ldots, y_m \in \mathbb{R}$ given outputs.

   (a) 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 $K(x,t) = \phi(x)^\top \phi(t)$.
   [5 *marks*]

   (b) Download the data set "atml.mat" from `http://www.cs.ucl.ac.uk/staff/a.argyriou/courses/index.html`. It consists of a training set of inputs (X) and outputs (y) and a test set of inputs (Xt) and outputs (yt). The dimensionality of the inputs is 10 and each column of the matrices X, Xt corresponds to one input x_i. Thus, there are 100 training data and 50 test data. Write and hand in Matlab code that performs ridge regression on this training set and computes the mean squared error on the test set. Use a regularisation parameter $\gamma = 10^{-6}$ and the following Gaussian kernels:

$$K(x,t) = e^{-\omega \|x-t\|^2}$$

   where $\omega = 10^s, s \in [-15, 1]$ (e.g. you can select a number of equally spaced values in this interval). Also plot the test error versus ω (use log-scales for both axes). Finally, plot the function $f : \mathbb{R}^{10} \to \mathbb{R}$ learned, by setting all coordinates of x apart from one to zero (that is, a cross-section of the function graph on one of the 10 coordinate axes). Note that you can use the "\" function in Matlab to solve a linear system.
   [15 *marks*]

   (c) 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, $\sigma_1, \ldots, \sigma_r$ are the singular values of W and r is the rank of W.
   [5 *marks*]

   [*Total* 25 *marks*]

2. This question pertains to convex functions and convex optimisation.

   (a) 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.
[10 *marks*]

(b) 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).
[10 *marks*]

(c) Solve the following linear program

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

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

[*Total* 25 *marks*]