58147 Machine Learning (Spring 2005)

Exercise 9 (Wednesday 6 April)

1. (a) Complete the proof of Theorem 4.7 by showing that if k_1 and k_2 are valid kernels, then so are $k_1 + k_2$ and ak_1 for $a > 0$.

   (b) Assume that k is a valid kernel with feature map $\boldsymbol{\psi}$. Obtain $\tilde{k}$ by *normalising*:

   $$\tilde{k}(x, z) = \frac{k(x, z)}{(k(x, x)k(z, z))^{1/2}}.$$

   Show that $\tilde{k}$ is a kernel for the feature map given by $\tilde{\boldsymbol{\psi}}(x) = \boldsymbol{\psi}(x) / \|\boldsymbol{\psi}(x)\|$.

2. Prove Corollary 4.8.

   Hints: For part 1, use induction. For part 2, notice that $\exp(r)$ is a limit of polynomials $p_n(r)$. For part 3, normalise the kernel $\exp(\langle x, z \rangle / \sigma^2)$.

3. Define $f(x) = x^2$ and $g(x) = (x - 2)^2$. Consider the problem

   $$\text{minimise} \quad f(x) \quad \text{subject to} \quad g(x) \leq 1.$$

   Write out the Lagrangian $L(x, \lambda)$. Solve the problem using the KKT conditions. Write out also the dual $g(\lambda)$ in a closed form.

   Use some mathematical software (*e.g.*, Maple) to plot $L(x, \lambda)$ (as a function of two variables, so you get a three-dimensional picture). Find from this plot the functions $x \mapsto \max_\lambda L(x, \lambda)$ and $\lambda \mapsto \min_x L(x, \lambda)$. Notice the point where they coincide.

   Repeat the exercise with the constraint $g(x) \leq 8$.

4. In an online prediction setting, given an example $(\boldsymbol{x}_t, y_t) \in \mathbf{R}^n \times \mathbf{R}$, let $W_t = \left\{ \boldsymbol{w} \in \mathbf{R}^n \mid (\boldsymbol{w} \cdot \boldsymbol{x}_t - y_t)^2 \leq R^2 \right\}$ where R is some constant. Then define the update as

   $$\boldsymbol{w}_{t+1} = \arg\min_{\boldsymbol{w} \in W_t} \|\boldsymbol{w}_t - \boldsymbol{w}\|_2^2.$$

   Write out the update in a closed form (*i.e.*, solve the minimisation).

5. Recall the relative entropy

   $$d_{\mathrm{re}}(\boldsymbol{p}, \boldsymbol{q}) = \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i}$$

   where we assume $p_i, q_i \geq 0$ and $\sum_{i=1}^n p_i = \sum_{i=1}^n q_i = 1$. Fix $\boldsymbol{q}$ such that $q_i > 0$ for all i. Using the KKT conditions, show that subject to the constraints $p_i \geq 0$, $i = 1, \ldots, n$, and $\sum_{i=1}^n p_i = 1$, the relative entropy $d_{re}(\boldsymbol{p}, \boldsymbol{q})$ is minimised at $\boldsymbol{p} = \boldsymbol{q}$. You may take it as known that $d_{\mathrm{re}}(\boldsymbol{p}, \boldsymbol{q})$ is convex in $\boldsymbol{p}$ (and also $\boldsymbol{q}$, although that is not needed here.)

   What is the dual?