GI01/4C55 – Homework #3 (Due 12am, November 23, 2005)

Aim: To get familiarity with kernels, regularization and VC–dimension. Presentation, clarity, and synthesis of exposition will be taken into account in the assessment of these exercises.

This document is available at `http://www.cs.ucl.ac.uk/staff/M.Pontil/courses/SL-homework3.pdf`

1. **[60 pts]** `basic` (*kernel properties*)

   Are the following functions $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ valid kernels? Explain your observation. When K is a valid kernel provide a feature map representation for it.

   (a) $K(\mathbf{x}, \mathbf{t}) = f(\mathbf{x})f(\mathbf{t})$, where $f : \mathbb{R}^d \to \mathbb{R}$.

   (b) $K(\mathbf{x}, \mathbf{t}) = f(\mathbf{x})g(\mathbf{t})$, where f and g are functions from $\mathbb{R}^d$ to $\mathbb{R}$.

   (c) $K(\mathbf{x}, \mathbf{t}) = \mathbf{x}^\top \mathbf{D}\mathbf{t}$, where $\mathbf{D}$ is a diagonal matrix with non-negative elements.

   (d) $K(\mathbf{x}, \mathbf{t}) = (1 - \mathbf{x}^\top \mathbf{t})^2$.

   (e) $K(\mathbf{x}, \mathbf{t}) = \frac{\mathbf{x}^\top \mathbf{t}}{\|\mathbf{x}\|\|\mathbf{t}\|}$.

   (f) $K(\mathbf{x}, \mathbf{t}) = \prod_{i=1}^{d} x_i t_i$ (note: we used the notation x_i for the i–th component of the vector $\mathbf{x} \in \mathbb{R}^d$).

2. **[20 pts]** `medium` (*Ridge regression*)

   Consider the ridge regression problem: $(\mathbf{w}_\lambda, b_\lambda) = \mathrm{argmin}_{\mathbf{w},b} \left\{\sum_{i=1}^{m} (y_i - b - \mathbf{w}^\top \mathbf{x}_i)^2 + \lambda \mathbf{w}^\top \mathbf{w}\right\}$. Show that this problem is equivalent to the problem

   $$(\hat{\mathbf{w}}_\lambda, \hat{b}_\lambda) = \mathrm{argmin}_{\mathbf{w},b} \left\{ \sum_{i=1}^{m} \left(y_i - b - \mathbf{w}^\top (\mathbf{x}_i - \bar{\mathbf{x}})\right)^2 + \lambda \mathbf{w}^\top \mathbf{w} \right\}$$

   where $\bar{\mathbf{x}}$ is the average of the input data. Give the correspondence between $(\hat{\mathbf{w}}_\lambda, \hat{b}_\lambda)$ and $(\mathbf{w}_\lambda, b_\lambda)$. Characterize the solution to the modified problem.

3. **[20 pts]** `medium/advanced` (*VC dimension of rectangles*)

   Consider the set $\mathcal{H}$ of classifiers generated by axis parallel rectangles in the plane (that is the set of intervals $[a_1, a_2] \times [b_1, b_2]$, where $a_2 \geq a_1$, $b_2 \geq b_1$) where each rectangle generates two classifiers, one which classifies the inside part as positive and the outside part as negative and the opposite classifier.

   Show that the VC–dimension of $\mathcal{H}$ equals 4. Does the result change if you remove the axis parallel assumption? Explain your observation.

4. **[10pts BONUS]**

   Prove that the VC–dimension of the set of classifiers $\{\mathrm{sign}(\sin(ax)) : a \in \mathbb{R}\}$ is infinite.