GI01/4C55 – Homework #1 (Due 12am, October 21, 2005)

Aim: To get familiarity with linear regression and least squares. 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-homework1.pdf`

1. **[60 pts]** `basic` (*least squares*)

   (a) Consider the constant function model $f(\mathbf{x}; b) = b$, where $b \in \mathbb{R}$. Derive the least squares solution for this model, that is, compute a minimizer of the empirical error

   $$\mathcal{E}(b) := \sum_{i=1}^{m} \big(y_i - f(\mathbf{x}_i; b)\big)^2.$$

   Does the solution depend on the dimension d of $\mathbf{x}$? Explain your observation.

   (b) Now, consider a linear model in one dimension: $f(x) = ax + b$, where $a, b \in \mathbb{R}$. Derive the least square solution for this model.

2. **[20 pts]** `medium` (*non linear least squares*)

   Suppose that the regression function $f^*(\mathbf{x})$ is quadratic in $\mathbf{x}$ and goes through the origin, that is, $f^*(\mathbf{0}) = 0$ (here, $\mathbf{0}$ is the d–dimensional vector all of whose components are equal to zero). Which hypothesis space of models would you choose in this case? Explain why this would be a good choice. Derive the formula of your learning algorithm on the training set S.

3. **[20 pts]** `medium` (*decomposition formula for the expected error*)

   Let $d = 1$ and define the expected error of a function $f : \mathbb{R} \to \mathbb{R}$ as

   $$\mathcal{E}(f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \big(y - f(x)\big)^2 dP(x, y).$$

   Derive the following formula:

   $$\mathcal{E}(f) = \mathcal{E}(f^*) + \int_{-\infty}^{\infty} \big(f^*(x) - f(x)\big)^2 dP(x).$$

   Explain the meaning of this formula and argue that it also holds true in the general case that $\mathbf{x} \in \mathbb{R}^d$ and $dP(\mathbf{x}, y)$ is an arbitrary probability measure on $\mathbb{R}^d \times \mathbb{R}$.

4. **[10pts BONUS]**

   Give an example of a probability density $P(\mathbf{x}, y)$ which does not satisfy the additive noise model and compute the underlying regression function. Explain your observation.

Remarks: In all questions, $S = \{(\mathbf{x}_i, y_i)\}_{i=1}^m \subseteq \mathbb{R}^d \times \mathbb{R}$ denotes a training set and $\mathbb{R} = \{-\infty, \infty\}$ is the set of real numbers. It is always assumed that S is generated *i.i.d.* from a probability measure $P(\mathbf{x}, y)$. The function f^* denotes the regression function, namely, $f^*(\mathbf{x}) = \int_{-\infty}^{\infty} y dP(y|\mathbf{x}) \equiv \mathbf{E}_{y|\mathbf{x}}[y]$. If you find it easier you may also assume that the probability measure $dP(\mathbf{x}, y)$ has a density, in which case you can write $dP(\mathbf{x}, y) = P(\mathbf{x}, y) d\mathbf{x} dy$. For example, in the additive gaussian noise model $P(\mathbf{x}, y) = \exp(-\beta(y - f^*(\mathbf{x}))^2) P(\mathbf{x})$.