# GI01/4C55 – Homework #2 (Due 12am, November 4, 2005)

**Aim:** To get familiarity with least squares, logistic regression and linear discriminant analysis. Presentation, clarity, and synthesis of exposition will be taken into account in the assessment of these exercises.

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, $\mathbf{X}$ is the $m \times d$ matrix $[\mathbf{x}_1, \ldots, \mathbf{x}_m]^\top$ and $\mathbf{y} = (y_1, \ldots, y_m)^\top$.

This document is available at `http://www.cs.ucl.ac.uk/staff/M.Pontil/courses/SL-homework2.pdf`

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

   Consider the weighted least square error

$$E(\mathbf{w}) = \sum_{i=1}^m d_i (y_i - \mathbf{w}^\top \mathbf{x}_i)^2$$

   where $d_i > 0$ is the weight assigned to point $i$.

   (a) Derive the modified normal equations for this model (that is, derive the equations $\nabla E(\mathbf{w}) = 0$).

   (b) Show that this model is justified via Maximum Likelihood (in a way similar to what we have seen in class for standard least squares) by the following modified additive noise model

$$P(y_i|\mathbf{x}_i; \mathbf{w}) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(y_i - \mathbf{w}^\top \mathbf{x}_i)^2}{2\sigma_i^2}\right)$$

   where $\sigma_i^2 = \frac{1}{d_i}$.

   (c) Show that $E(\mathbf{w})$ can be rewritten as $E(\mathbf{w}) = (\mathbf{X}\mathbf{w} - \mathbf{y})^\top \mathbf{D}(\mathbf{X}\mathbf{w} - \mathbf{y})$ for an appropriate choice of the matrix $\mathbf{D}$. Explain what this matrix is.

2. **[20 pts]** `medium` (*LDA and logistic regression*)

   Consider the linear discriminant analysis model for binary classification, where the class probability densities, $P(\mathbf{x}|0)$ and $P(\mathbf{x}|1)$, are assumed to be Gaussians with different means $\mu_0$ and $\mu_1$ and same covariance $\Sigma$. Denote the prior probability of class '0' by $\pi_0$. Show that the class conditional probability $P(y = 1|\mathbf{x})$ has the same form as in logistic regression, namely

$$P(y = 1|\mathbf{x}) = \frac{1}{1 + e^{-(\mathbf{w}^\top \mathbf{x} + b)}}$$

   and describe how the parameters $\mathbf{w}$ and $b$ depend on the parameters $\mu_0, \mu_1, \Sigma, \pi_0$.

3. **[20 pts]** `medium/advanced` (*logistic regression*)

   Consider the logistic regression error criterion

$$E(\mathbf{w}) = -\sum_{i=1}^m y_i \log p(\mathbf{w}^\top \mathbf{x}_i) + (1 - y_i) \log\left(1 - p(\mathbf{w}^\top \mathbf{x}_i)\right)$$

   where $p(z) = (1 + e^{-z})^{-1}$.

   Compute the Hessian of $E$, that is, the matrix-valued function $\mathbf{H}$ with elements

$$H_{jk} = \frac{\partial^2 E(\mathbf{w})}{\partial w_j \partial w_k}, \quad j, k = 1, \ldots d.$$

   Show that $\mathbf{H}$ is positive semi-definite, that is, for every $\mathbf{u} \in \mathbb{R}^d$, it holds that $\mathbf{u}^\top \mathbf{H}(\mathbf{w})\mathbf{u} \geq 0$.

   **Note:** Since, in general, a function $E$ is convex if and only if its Hessian is positive semidefinite (for every $\mathbf{w}$), this exercise shows that the logistic regression error criterion is convex.

   **[5 pts BONUS]** Can you think of a simpler way of showing that $E$ is convex?