Data Mining Coursework

Brad Baxter

December 23, 2003

My apologies for the late appearance of this coursework. I have reduced the quantity of work in consequence. It is due at the end of January and should be given in at our General Office on the seventh floor. Alternatively, you can use the wooden box outside the General Office.

1. We have seen in lectures that

$$r = \int_0^\infty \left(1 - e^{-r^2 t}\right) w(t)\, dt, \qquad r \ge 0,$$

where $w(t) = (4\pi)^{-1/2} t^{-3/2}$. Hence derive

$$(s^2 + c^2)^{1/2} = \int_0^\infty \left(1 - e^{-(s^2+c^2)t}\right) w(t)\, dt, \qquad s \ge 0,$$

and show that

$$(s^2 + c^2)^{1/2} = c + \int_0^\infty \left(1 - e^{-s^2 t}\right) \alpha(t)\, dt, \qquad s \ge 0,$$

where $\alpha(t)$ is to be determined. Use this integral relation to prove that

$$\sum_{j=1}^n \sum_{k=1}^n y_j y_k \sqrt{\|\mathbf{x}_j - \mathbf{x}_k\|^2 + c^2} \le 0$$

when $\sum_{\ell=1}^n y_\ell = 0$, for any vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ in $\mathbb{R}^d$.

2. The *Frobenius norm* of an $n \times n$ matrix A is defined by

$$\|A\|_F = \left(\sum_{j=1}^n \sum_{k=1}^n A_{jk}^2 \right)^{1/2}.$$

In other words, if we think of A as a long vector with n^2 elements, then $\|A\|_F$ is just the Euclidean norm of that vector. You can compute the Frobenius norm in Matlab using the command `norm(A, 'fro')`.

Several applications, such as robotics and aircraft control systems, present the following problem: the system attempts to maintain an orthogonal matrix $Q(t)$ that describes its orientation at each time t (the columns of the matrix are orthonormal vectors fixed in the system). Unfortunately, measurement errors occur which cause the measured $Q(t)$ to lose orthogonality, i.e. we no longer have $Q(t)^T Q(t) = I$. Therefore there is a need to calculate an orthogonal matrix U that is closest to the observed matrix $Q(t)$ in some sense. If we decide to choose U to minimize the Frobenius norm $\|Q(t) - U\|_F$, then there is a clever algorithm for calculating U: we choose $V_0 = Q(t)$ and then set

$$V_{\ell+1} = \frac{1}{2}\Big(V_\ell + (V_\ell^{-1})^T\Big), \qquad \ell \geq 0.$$

It can be shown that $\|V_\ell - U\|_F \to 0$ as $\ell \to \infty$, and you will see that convergence is fast (in most cases 4 steps will be enough). Write a short Matlab script to generate the matrices of this iteration and investigate the speed of convergence (for 3×3 matrices) by plotting $\log \|U - V_\ell\|_F$. You will need to generate random orthogonal matrices, for which the following Matlab code is suitable.

```
A = randn(3);
[Q, R] = qr(A);
```

You can assume that Q is a suitable random orthogonal matrix. You can then slightly perturb Q by setting

```
V = Q + delta*randn(3);
```

Of course, a large `delta` is a large perturbation. I suggest starting with `delta = 0.1`, but try larger values also. Does the algorithm ever fail? Generate some histograms displaying the average behaviour for fixed delta and many random initial perturbed orthogonal matrices. [This problem is closely related to the Procrustes' problem, which may have been mentioned in one of our other courses. You can enhance your Christmas by discovering Procrustes' sadistic practices via Google.]

3. One way to use radial basis functions in regression methods is as follows. We are given sequences of points $\mathbf{b}_1, \ldots, \mathbf{b}_m$ and $\mathbf{c}_1, \ldots, \mathbf{c}_n$ lying in $\mathbb{R}^d$ and, given function values $f_1, \ldots, f_n$, we seek real coefficients $a_1, \ldots, a_m$ minimizing the sum of squares

$$\sum_{\ell=1}^{n} \left(f_\ell - s(\mathbf{c}_\ell)\right)^2,$$

where

$$s(\mathbf{x}) = \sum_{k=1}^{n} a_k \phi(\mathbf{x} - \mathbf{b}_k),$$

for some radially symmetric function $\phi : \mathbb{R}^d \to \mathbb{R}$. One solution to this problem, as for any least squares problem, is to solve the *normal equations*, which are given by

$$A^T A\mathbf{a} = A^T \mathbf{f},$$

where $\mathbf{a} = (a_1, \ldots, a_m)^T$, $\mathbf{f} = (f_1, \ldots, f_n)^T$ and

$$A_{\ell k} = \phi(\mathbf{c}_\ell - \mathbf{b}_k), \qquad 1 \le k \le m, \quad 1 \le \ell \le n.$$

Show that

$$(A^T A)_{jk} = \sum_{\ell=1}^{n} \phi(\mathbf{c}_\ell - \mathbf{b}_j)\phi(\mathbf{c}_\ell - \mathbf{b}_k), \qquad 1 \le j, k \le m,$$

and hence derive

$$\mathbf{v}^T A^T A \mathbf{v} = \sum_{\ell=1}^{n} \left(\sum_{k=1}^{m} v_k \phi(\mathbf{c}_\ell - \mathbf{b}_k) \right)^2 \ge 0.$$

Now suppose that $\phi(\mathbf{x}) = e^{-\lambda \|\mathbf{x}\|^2}$, for $\mathbf{x} \in \mathbb{R}^d$, λ being a positive constant. Further, suppose that $\mathbf{b}_k^T \mathbf{c}_\ell = 0$, for all k and ℓ. Prove that there is a nonzero vector $\mathbf{v}$ for which $\mathbf{v}^T A^T A \mathbf{v} = 0$. Thus the matrix for the normal equations can be singular in some special cases (I discovered this new result last summer).