Data Mining Examination Solutions

Brad Baxter

January 13, 2006

1. (i). The matrices $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal matrices, whilst $S \in \mathbb{R}^{m\times n}$ is a diagonal matrix whose diagonal elements satisfy $s_1 \ge s_2 \ge \cdots \ge s_n \ge 0$. These diagonal elements of S are called the singular values of A.

4 pts

(ii). Pre- and post-multiplying by orthogonal matrices leaves the Euclidean norm of vectors invariant. Thus

$$\|A\mathbf{x} - \mathbf{y}\|^2 = \|USV^T\mathbf{x} - \mathbf{y}\|^2 = \|S\mathbf{a} - U^T y\|^2 = \|S\mathbf{a} - \mathbf{b}\|^2.$$

4 pts

(iii). Using the previous part of the question,

$$\|A\mathbf{x} - \mathbf{y}\|^2 = \sum_{k=1}^{n} (s_k a_k - b_k)^2 + \sum_{\ell > n} b_\ell^2 \ge \sum_{\ell > n} b_\ell^2.$$

If every singular value of A is positive, then we can attain this lower bound by setting $a_k = s_k^{-1} b_k$, for $1 \le k \le n$. Thus

$$\mathbf{a} = T\mathbf{b},$$

or

$$V^T\mathbf{x}^* = TU^T\mathbf{y},$$

i.e.

$$\mathbf{x}^* = VTU^T\mathbf{y}.$$

4 pts

(iv). We need to show that

$$VTU^T = (A^TA)^{-1}A^T,$$

and I would be tempted to reward any student who remembered that this is called the Moore–Penrose pseudo-inverse of A. Now

$$\begin{aligned}
(A^TA)^{-1}A^T &= \left((USV^T)^T USV^T\right)^{-1} (USV^T)^T \\
&= \left(VS^TU^TUSV^T\right)^{-1} VS^TU^T \\
&= \left(VS^TSV^T\right)^{-1} VS^TU^T \\
&= V(S^TS)^{-1}V^TVS^TU^T \\
&= V(S^TS)^{-1}S^TU^T.
\end{aligned}$$

But $(S^TS)^{-1}$ is the inverse of the $n \times n$ diagonal matrix whose diagonal elements are $s_1^{-2}, \ldots, s_n^{-2}$. Hence

$$(S^TS)^{-1}S^T = T,$$

as required.

8 pts

2. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be points in $\mathbb{R}^d$. The k-means algorithm is a simple method for iteratively updating a set of k *cluster centres* $\mathbf{m}_1, \ldots, \mathbf{m}_k$. At the start of the algorithm, these points can be any vectors.

   Now the k cluster centres partition $\mathbb{R}^d$ into k clusters: we let the ith cluster C_i be those points in $\mathbb{R}^d$ for which $\mathbf{m}_i$ is the closest cluster centre, that is

$$C_i = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x} - \mathbf{m}_i\| = \min_{1 \le \ell \le k} \|\mathbf{x} - \mathbf{m}_\ell\|\}, \qquad 1 \le i \le n,$$

   and students are not expected to deal with ambiguous cases for which some points lie in more than one cluster. We then replace each cluster centre $\mathbf{m}_i$ by the centroid of the subset of points in $\mathbf{x}_1, \ldots, \mathbf{x}_n$ which are contained in the ith-cluster (the centroid of a finite set of points $\mathbf{v}_1, \ldots, \mathbf{v}_j$ is simply the sample average $(\mathbf{v}_1 + \cdots + \mathbf{v}_j)/j$). The new cluster centres then define corresponding new centres, and we then repeat the procedure until the cluster centres converge.

8 pts

(i). Here $(-R, \pm 1)$ lie in the left cluster, whilst $(R, \pm 1)$ lie in the right cluster. Hence the new centroids are $(-R, 0)$ and $(R, 0)$, and no further change occurs.

4pts

(ii). Here $(\pm R, 1)$ lie in the upper cluster, while $(\pm R, -1)$ lie in the lower cluster. Hence the new centroids are $(0, -1)$ and $(0, 1)$, and no further change occurs.

4pts

(iii). If $(R, \pm 1)$ lie in the same cluster, that is, $u^\perp$ separates $(R, \pm 1)$ and $(-R, \pm 1)$, then the new centroids are $(\pm R, 0)$, and no further progress occurs.

However, if $(\pm R, 1)$ lie in the same cluster, that is, $u^\perp$ separates $(\pm R, 1)$ and $(\pm R, -1)$, then the new centroids are $(0, \pm 1)$, and no further progress occurs.

4 pts

3. (i). We have

$$\begin{aligned}\Gamma(n+1/2) &= \int_0^\infty e^{-t}t^{n-1/2}\,dt \\ &= \int_0^\infty e^{-a^2 s}a^{2n-1}s^{n-1/2}a^2\,ds \\ &= a^{2n+1}\int_0^\infty e^{-a^2 s}s^{n-1/2}\,ds.\end{aligned}$$

6 pts

(ii). Setting $a^2 = r^2 + c^2$ in the previous integral, we obtain

$$\begin{aligned}(r^2+c^2)^{-(2n+1)/2} &= \int_0^\infty e^{-(r^2+c^2)s}s^{n-1/2}\,ds \\ &= \int_0^\infty e^{-r^2 s}e^{-c^2 s}s^{n-1/2}\,ds.\end{aligned}$$

6 pts

(iii). Using the integral derived in the second part of this question, we find

$$\sum_{j=1}^n\sum_{k=1}^n v_k v_k(\|\mathbf{x}_j-\mathbf{x}_k\|^2+c^2)^{-(2n+1)/2} = \int_0^\infty\left(\sum_{j=1}^n\sum_{k=1}^n v_k v_k\exp(-s\|\mathbf{x}_j-\mathbf{x}_k\|^2)\right)w(s)\,ds.$$

The integrand is strictly positive for all $s > 0$ if the points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are distinct, because the Gaussian is a strictly positive definite function (given in the question). Thus we have shown that the interpolation matrix for this radial basis function is also strictly positive definite.

8 pts

4.

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

2 pts

If $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are the columns of A, then

$$\begin{aligned} \|UA\|_F^2 &= \|U\mathbf{a}_1\|^2 + \cdots + \|U\mathbf{a}_n\|^2 \\ &= \|\mathbf{a}_1\|^2 + \cdots + \|\mathbf{a}_n\|^2 \\ &= \|A\|_F^2. \end{aligned}$$

Thus pre-multiplication by an orthogonal matrix leaves the Frobenius norm unchanged. For post-multiplication, we use the fact that $\|A\|_F = \|A^T\|_F$, so that

$$\|AV\|_F = \|(AV)^T\|_F = \|V^T A^T\|_F = \|A^T\|_F = \|A\|_F.$$

6 pts

(i). If $A = USV^T$ is the SVD of A, then $Q_A = UV^T$.

4 pts

(ii). If $B^T A = USV^T$ is the SVD of $B^T A$, then $\hat{Q} = UV^T$.

4 pts

(iii). If $A = USV^T$ is the SVD of A, then we let

$$S_r = \text{diag}\,\{s_1, \ldots, s_r, 0, \ldots, 0\} \in \mathbb{R}^{m \times n}$$

and define $A_r = US_rV^T$.

4 pts