Statistics 640
Spring, 2009

Data Mining and Statistical Learning

Professor David W. Scott
2057 Duncan Hall, 713-348-6037, scottdw@rice.edu

Tu/Th 2:30-3:50 Keck Hall - 101
Office Hours: MW 1:30-2:30 or by appt.

First Assignment: Due in class January 13, 2009

Download the useful visualization software called *ggobi* from the web (find it) and install it on your PC or unix box. Also install the R language if you have not done so already. Grab the *zipcode data* from the course web site

www.stat.rice.edu/∼scottdw/stat640/zip.code/

Each line of this file contains a single number between 0 and 9 followed by a vector of length 256 which contains a 16 by 16 grayscale image of a handwritten digit of that number. Choose any two digits, and extract all of the examples in the data file. Use the ggobi software to "visualize" these 256-dimensional vectors. In particular, try to determine if the data clouds for your 2 numbers "separate" in the high-dimensional space.

You may discuss your work (and collaborate) with others in the class, but turn in a short separate report of your findings. (Make sure you have a working version of ggobi of your own. There are interfaces between ggobi and R available as well.)

Multivariate Random Variable Background–For Class

1. vector random variable $\mathbf{X}$ $\quad p \times 1$

$$\mu = E\,\mathbf{X} = \begin{pmatrix} E\,X_1 \\ \vdots \\ E\,X_p \end{pmatrix}$$

$$\begin{aligned} \Sigma = \mathrm{Cov}(\mathbf{X}) =& \Big(\; E[(X_i - \mu_i)(X_j - \mu_j)] \; \Big) \\ =& E\left[(\mathbf{X} - \mu)(\mathbf{X} - \mu)^T\right] \qquad \text{(outer product)} \end{aligned}$$

2. $\mathbf{Y} = A\mathbf{X} + \mathbf{b}$ $\quad q \times 1$ $\quad A$ is $q \times p$

$$E\,\mathbf{Y} = A\,E\,\mathbf{X} + \mathbf{b} = A\mu_X + \mathbf{b}$$

$$\begin{aligned} \mathrm{Cov}(\mathbf{Y}) &= E(\mathbf{Y}-\mu_Y)(\mathbf{Y}-\mu_Y)^T \qquad \mathbf{Y}-\mu_Y = A\mathbf{X}+\mathbf{b}-(A\mu_X+\mathbf{b}) \\ &= E\left[A(\mathbf{X}-\mu_X)(\mathbf{X}-\mu_X)^T A^T\right] \\ &= A\Sigma_X A^T \qquad q \times q \quad \text{by linearity} \end{aligned}$$

3. $MN(\mu, \Sigma)$ $\quad f(\mathbf{x}) = |2\pi\Sigma|^{-1/2}\exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\right)$ $\quad$ or $\quad (2\pi)^{-p/2}|\Sigma|^{-1/2}$

4. Spectral decomposition $\quad \Sigma\mathbf{e}_i = \lambda_i\mathbf{e}_i \quad \mathbf{e}_i \perp \mathbf{e}_j \iff \mathbf{e}_i^T\mathbf{e}_j = \delta_{ij}$

Note: $\quad \mathbf{a}^T\Sigma\mathbf{a} = E\left[\mathbf{a}^T(\mathbf{X}-\mu)(\mathbf{X}-\mu)^T\mathbf{a}\right] = E\left[\mathbf{a}^T(\mathbf{X}-\mu)\right]^2 \geq 0$
therefore $\mathbf{e}_i^T\Sigma\mathbf{e}_i = \lambda_i \geq 0$ and Σ is p.s.d. $\quad$ (pos def $\iff \lambda_i > 0$)

Matrix form: $\quad \Sigma\mathbf{E} = \mathbf{E}\Lambda \quad$ or $\quad \Sigma = \mathbf{E}\Lambda\mathbf{E}^T \quad$ where $\mathbf{E} = \left[\mathbf{e}_1\,\mathbf{e}_2\,\ldots\,\mathbf{e}_p\right]$

5. $\mathbf{X} \sim MN(\mu_X, \Sigma_X)\,, \quad \mathbf{Y} = A\mathbf{X} + \mathbf{b} \sim MN(A\mu_X + \mathbf{b}, A\Sigma_X A^T)$

6. Assuming $\lambda_i > 0$,

$$\begin{aligned} \Sigma &= \mathbf{E}\Lambda\mathbf{E}^t & \Sigma^{-1} &= \mathbf{E}\Lambda^{-1}\mathbf{E}^T \\ \Sigma^{1/2} &= \mathbf{E}\Lambda^{1/2}\mathbf{E}^t & \Sigma^{-1/2} &= \mathbf{E}\Lambda^{-1/2}\mathbf{E}^T \end{aligned}$$

7. Sphering data

$$Y = \Sigma_X^{-1/2}\mathbf{X} \qquad \Longrightarrow \qquad \Sigma_Y = \Sigma_X^{-1/2}\Sigma_X\Sigma_X^{-/2} = I_p$$

therefore $\quad \mathbf{X} \sim MN(\mu_X, \Sigma_X) \quad \Longrightarrow \quad Y = \Sigma_X^{-1/2}(\mathbf{X}-\mu_X) \sim MN(\mathbf{0}, I_p)$

8. Vector and matrix differentiation (for a symmetric matrix, A)

$$\begin{aligned} g(\mathbf{x}) &= \mathbf{a}^T\mathbf{x} = \mathbf{x}^T\mathbf{a} & \nabla_x g(\mathbf{x}) &= \mathbf{a} \\ h(\mathbf{x}) &= \mathbf{x}^T A\mathbf{x} & \nabla_x h(\mathbf{x}) &= 2A\mathbf{x} \end{aligned}$$

Finally,

$$\nabla_x\nabla_x h(\mathbf{x}) = \nabla_x^2 h(\mathbf{x}) = 2A$$

which is sometimes written this way:

$$\nabla_x\nabla_x^T h(\mathbf{x}) = \nabla_x^2 h(\mathbf{x}) = 2A\;.$$

Reference: K.V. Mardia, J.T. Kent, and J.M. Bibby (1979). *Multivariate Analysis*, Academic Press, London.