Multivariate Analysis

Homework #1 (2024-2025), Answers

Q1. Derivatives of traces of matrix products

This is useful for solving the minimization equations for regression and related procedures. Let A be an $M \times N$ matrix, U an $N \times M$ matrix, V an $N \times M$ matrix, W an $M \times M$ matrix, and Z an $N \times N$ matrix. Assume all entries are real, so that the adjoint is equal to the transpose.

Compute the partial derivatives with respect to each element $a_{j,k}$ of each of the following, and assemble into a matrix: $tr(UA)$, $tr(AV)$, $tr(A^*WA)$, and $tr(AZA^*)$.

$tr(UA) = \sum_{m,n} u_{n,m} a_{m,n}$ so $\frac{\partial}{\partial a_{j,k}} tr(UA) = \frac{\partial}{\partial a_{j,k}} \sum_{m,n} u_{n,m} a_{m,n} = u_{k,j}$, which we can write as $\frac{d}{dA} tr(UA) = U^*$.

$tr(AV) = \sum_{m,n} a_{m,n} v_{n,m}$ so $\frac{\partial}{\partial a_{j,k}} tr(AV) = \frac{\partial}{\partial a_{j,k}} \sum_{m,n} a_{m,n} v_{n,m} = v_{k,j}$, which we can write as $\frac{d}{dA} tr(AV) = V^*$.
Alternatively, $tr(AV) = tr(VA)$, so this result follows from $\frac{d}{dA} tr(UA) = U^*$.

$tr(A^*WA) = \sum_{m,n,p} a_{n,m} w_{n,p} a_{p,n}$ so $\frac{\partial}{\partial a_{j,k}} tr(A^*WA) = \frac{\partial}{\partial a_{j,k}} \sum_{m,n,p} a_{n,m} w_{n,p} a_{p,m} = \sum_{p} w_{j,p} a_{p,k} + \sum_{n} a_{n,k} w_{n,j}$ (the last step using the product rule for derivatives), which we can write as $\frac{d}{dA} tr(A^*WA) = WA + W^*A = (W + W^*)A$.

$tr(AZA^*) = \sum_{m,n,p} a_{m,n} z_{n,p} a_{m,p}$ so $\frac{\partial}{\partial a_{j,k}} tr(AZA^*) = \frac{\partial}{\partial a_{j,k}} \sum_{m,n,p} a_{m,n} z_{n,p} a_{m,p} = \sum_{p} z_{k,p} a_{j,p} + \sum_{n} a_{j,n} z_{n,k}$ (the last step using the product rule for derivatives), which we can write as $\frac{d}{dA} tr(AZA^*) = AZ^* + AZ = A(Z + Z^*)$.

Q2: Standard regression

Setup for standard regression: You are given a matrix of measured data, Y, of n rows and k columns: each row is a measurement, univariate if $k = 1$ and multivariate of $k > 1$. You are also given a set of p regressors, a matrix X, of n rows and p columns. The goal is to fit the measured data with $Y^{fit} = XB$, where B, the matrix to be found, is the set of regression coefficients. Standard regression determines B in the model $Y^{fit} = XB$ by minimizing $R_{std}^2 = \mathrm{tr}\left((Y - Y^{fit})^*(Y - Y^{fit})\right)$.

Use the results of Q1 and Q2 to find B by setting $\frac{d}{dB} R_{std}^2 = 0$.

$$\frac{d}{dB}R_{std}^2 = \frac{d}{dB}\left[tr\left((Y-Y^{fit})^*(Y-Y^{fit})\right)\right] = \frac{d}{dB}\left[tr\left((Y-XB)^*(Y-XB)\right)\right]$$
$$= \frac{d}{dB}\mathrm{tr}\left(-B^*X^*Y - Y^*XB + B^*X^*XB + Y^*Y\right) = \frac{d}{dB}\left[-\mathrm{tr}\left(B^*X^*Y\right) - tr\left(Y^*XB\right) + tr\left(B^*X^*XB\right)\right],$$
$$= \frac{d}{dB}\left[-\mathrm{tr}\left(Y^*XB\right) - tr\left(Y^*XB\right) + tr\left(B^*X^*XB\right)\right] = \frac{d}{dB}\left[-2\,\mathrm{tr}\left(Y^*XB\right) + tr\left(B^*X^*XB\right)\right]$$
where we have used only that Y is constant, and that $tr(M^*) = tr(M)$ for a real square matrix. Then, from Q1: $\frac{d}{dB}\mathrm{tr}\left(Y^*XB\right) = \left(Y^*X\right)^* = X^*Y$, $\frac{d}{dB}tr\left(B^*X^*XB\right) = \left(X^*X + \left(X^*X\right)^*\right)B = 2X^*XB$, and $\frac{d}{dB}tr\left(B^*B\right) = 2B$.

So
$$\frac{d}{dB}R_{std}^2 = \frac{d}{dB}\left[-2\,\mathrm{tr}\left(Y^*XB\right) + tr\left(B^*X^*XB\right)\right] = -2X^*Y + 2X^*XB = -2X^*Y + 2(X^*X)B \ ,$$
and
$$\frac{d}{dB}R_{std}^2 = 0 \Leftrightarrow (X^*X)B = X^*Y \Leftrightarrow B = (X^*X)^{-1}X^*Y$$

Q3: Ridge regression

Ridge regression uses the same set-up as standard regression, but puts a premium on finding models in which the regression coefficients are small (in addition to how well they fit the data). The extent to which the criterion of "small regression coefficients" is weighted is given by a parameter k. More formally, ridge regression determines B by minimizing $R_{ridge}^2 = R_{std}^2 + k\bullet tr(B^*B) = tr\left((Y-Y^{fit})^*(Y-Y^{fit})\right) + k\bullet tr(B^*B)$. Use the results of Q1 to find B by setting $\frac{d}{dB}R_{ridge}^2 = 0$.

Since $R_{ridge}^2 = R_{std}^2 + k\bullet tr(B^*B)$, and using Q2 for $\frac{d}{dB}R_{std}^2$,
$$\frac{d}{dB}R_{ridge}^2 = \frac{d}{dB}R_{std}^2 + \frac{d}{dB}\left[k\bullet tr(B^*B)\right] = -2X^*Y + 2(X^*X)B + \frac{d}{dB}\left[k\bullet trB^*B\right] = -2X^*Y + 2(X^*X + kI)B \ ,$$
where we have used only that Y is constant, and that $tr(M^*) = tr(M)$ for a real square matrix. Then
$$\frac{d}{dB}R_{ridge}^2 = 0 \Leftrightarrow (X^*X + kI)B = X^*Y \Leftrightarrow B = (X^*X + kI)^{-1}X^*Y$$