Math Camp - Day 5

Aaron Rudkin

Trinity College Dublin
Department of Political Science

rudkina@tcd.ie

Inverse of a Matrix

- The inverse of a matrix A is the matrix A^{-1} that, when multiplied with A produces a $n \times n$ identity matrix
  $AA^{-1} = A^{-1}A = I_{n \times n}$
- Only square matrices can have inverses
- Only matrices with $det(A) \neq 0$ are invertible.

Inverse of a Matrix

- Inverse operation also called the **solution** of the matrix.
- It is possible, though annoying, to find the inverse of a matrix. Don't bother. Use a computer.
- Generical intuition: Create theoretical inverse, use algebra to solve equation:

$$\begin{bmatrix} 1 & -2 \\ 2 & -3 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$1 = 1a - 2c$$

$$0 = 1b - 2d$$

$$0 = 2a - 3c$$

$$1 = 2b - 3d$$

Use system of equations to solve:

$$\begin{bmatrix} -3 & 2 \\ -2 & 1 \end{bmatrix}$$

System of Equations Recap

- Substitution: One equation for one variable, substitute into another equation and solve

System of Equations Recap

- Substitution: One equation for one variable, substitute into another equation and solve
  - The benefit of substitution is primarily that, as long as you do the algebra right, you'll find the answer, and you don't have to remember any procedure for doing so. The downside is it is not always very systematic, and you can waste a lot of time on algebra that you didn't need to do.
- Elimination: Subtract a multiple of one equation from another to eliminate a term

System of Equations Recap

- Substitution: One equation for one variable, substitute into another equation and solve
  - The benefit of substitution is primarily that, as long as you do the algebra right, you'll find the answer, and you don't have to remember any procedure for doing so. The downside is it is not always very systematic, and you can waste a lot of time on algebra that you didn't need to do.
- Elimination: Subtract a multiple of one equation from another to eliminate a term
- Matrix inversion: Treat vectors as columns of matrix, weights as a column vector, solve

System of Equations Recap

- Substitution: One equation for one variable, substitute into another equation and solve
  - The benefit of substitution is primarily that, as long as you do the algebra right, you'll find the answer, and you don't have to remember any procedure for doing so. The downside is it is not always very systematic, and you can waste a lot of time on algebra that you didn't need to do.
- Elimination: Subtract a multiple of one equation from another to eliminate a term
- Matrix inversion: Treat vectors as columns of matrix, weights as a column vector, solve
- Cramer's rule: Look up online; special property of determinants of matrices.

Matrix Inverse Properties

Inverse	$(A^{-1})^{-1} = A$
Multiplicative property	$(AB)^{-1} = B^{-1}A^{-1}$
Scalar multiplication ($n \times n$)	$(cA)^{-1} = c^{-1}A^{-1}$ if $c \neq 0$
Transpose	$(A^{-1})^T = (A^T)^{-1}$

Linear Independence

Linear Independence

- Given a set of vectors $v : v_1, v_2, v_3, \ldots v_n$, the vectors can be said to be **linearly dependent** (**collinear**) or **linearly independent**
- The vectors are linearly dependent if some combination of the vectors results in the zero vector and independent if no such combination exists
- Formally, the math is that vectors are linearly independent if no solution to $a_1 v_1 + a_2 v_2 + \ldots + a_n v_n = 0$ except for $a_1 = a_2 = \ldots = a_n = 0$
- Are $a = (1, 3, 5)$ and $b = (3, 9, 15)$ linearly independent? No. Setting the a_1 weight to -3 and the a_2 weight to 1 results in a 0 vector.
- Are $c = (2, 2, 0)$ and $d = (1, -1, 1)$ linearly independent? Yes. The third component requires a zero weight on the second vector, which then requires a zero weight on the first vector

Matrix Rank

Matrix Rank

- The **rank** of a matrix is defined as (a) the maximum number of linearly independent column vectors in the matrix or (b) the maximum number of linearly independent row vectors in the matrix. Both definitions are equivalent.
- The maximum rank of a matrix is the smaller of its two dimensions, for an $i \times j$ matrix;
    - If i is less than j, then the maximum rank of the matrix is i.
    - If i is greater than j, then the maximum rank of the matrix is j.
- Rank of A and B

$$A = \begin{pmatrix} 1 & 2 & 1 & 3 \\ 2 & 3 & 1 & 3 \end{pmatrix} \qquad B = \begin{pmatrix} 2 & 0 & 2 \\ 3 & 1 & 4 \\ 1 & 4 & 5 \\ 3 & 3 & 6 \end{pmatrix}$$

Multivariate Calculus

Functions of Several Variables

- We've seen functions of more than one variable before:

$$f(x, y, z)$$

- In the past we've talked about scalar outputs, where the answer is a number

$$f(x, y, z) = x^2 + xy^2 - z^3$$

- Vector outputs are also possible, with one equation for each output:

$$f(x, y, z) = \begin{pmatrix} x^2y - z \\ y^3z^2 - x^2 \\ \frac{3x^3z^2}{y^2} \end{pmatrix}$$

Partial Derivatives

- We ran into partial derivatives before now; when we take the derivative of a multivariable equation with respect to one of the variables.
- Say we have the function $f(x,z) = xz$. The partial derives respect the instantangeous rates of change in y due to the variable we differentiate, holding the other constant:

$$\frac{\delta y}{\delta x}xz = zdx \qquad \frac{\delta y}{\delta z}xz = xdz$$

- In other words, the partial derivative is linked intrinsically to "marginal effects", substantively. A marginal effect asks how a certain factor (education, salary, vaccination, training, audits) can effect an outcome of interest; but that outcome is also driven by other factors, and so we wish to "hold the other variables constant" while manipulating or observing variation in our variable of interest.

Gradient Vector

- **Gradient vector** is a special vector that includes all of the partial derivatives of a system of equations.
- $f(x, y, z) = 4x + 6y - z$
  - $\nabla f = \begin{pmatrix} 4 \\ 6 \\ -1 \end{pmatrix}$
- Solving optimization problems of multiple variables that do not have analytic solutions often requires using "gradient descent" – an effort to find values that lower the gradient, in order to find the extrema of inteest

Jacobian Matrix

- Jacobian matrix —**J**—is something of an extension; given i equations of k variables, the Jacobian is the matrix that contains every partial derivative of every equation
- Each row has the partial derivatives with respect to each variable of a single equation
- Each column has the partial derivatives with respect to one variable of every equation
- N_{row} = The number of functions and N_{column} =The number of unknowns

$$f_1 = x^2 + 3xy - z^3$$
$$f_2 = xy^2 - xz^2$$
$$f_3 = xyz - xy^4 + z^2$$
$$f_4 = x + y + z$$

$$J = \begin{pmatrix} 2x + 3y & 3x & 3z^2 \\ y^2 - z^2 & 2xy & -2xz \\ yz - y^4 & xz - 4xy^3 & 2z \\ 1 & 1 & 1 \end{pmatrix}$$

Hessian Matrix

- The **Hessian** matrix is a square matrix of every second-order partial derivatives of a scalar-valued function
- By second-order, we mean taking the derivative with respect to each set of two variables (e.g. for a two-variable system: (x, x), (x, y), (y, x), and (y, y))

$$\begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 x_n} \\ \frac{\partial^2 f}{\partial x_2 x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n x_1} & \frac{\partial^2 f}{\partial x_n x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

The Shape of Things to Come

- If you have not already, you will still encounter the magic of **OLS**, ordinary least squares, the hammer in your toolkit for fitting linear models.
- Given a vector of outcome variables y and a matrix of predictors X, OLS posits the following data generating mechanism:

  $$y = X\beta + \epsilon$$

- In other words, data can be explained by assigning weights β to each of the predictors in X, and the data also contains a stochastic component ϵ that cannot be explained.
- How do we choose the weights $\hat{\beta}$? We want to choose the best weights $\hat{\beta}$ (the hat implies our "best estimate") to minimize the sum of squared errors.

The Shape of Things to Come

- Given the error:

$$e = y - X\hat{\beta}$$

- Note we've swapped from ϵ (the underlying and unobservable, unexplaineable stochasticity in the data) to e, the portion of the variation unexplained by our model but observable.
- Rather than optimize the error, we optimize the sum of squared errors:

$$e'e = (y - X\hat{\beta})'(y - X\hat{beta})$$

- Notice that $e'e$ is a $(1 \times n) \times (n \times 1)$ matrix – in other words, a 1 by 1 scalar.
- Later, you'll learn that matrices and vectors can also do complex algebra:

$$e'e = y'y - 2\hat{\beta}'X'y - y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta}$$

- Here we've used FOIL.

The Shape of Things to Come

- Now we should take the derivative with respect to $\hat{\beta}$ to find the local minima – which values of beta minimize the sum of squared errors?

$$\frac{\delta}{\delta\hat{\beta}}e'e = -2X'y + 2X'X\hat{\beta} = 0$$

- Yes, matrices and vectors can use calculus, and so we can take their derivatives! A little tricky, but still... let's continue:

$$2X'X\hat{\beta} = 2X'y$$
$$X'X\hat{\beta} = X'y$$

- One final step: We want to isolate $\hat{\beta}$ on the left hand side, and we can do this by left-multiplying by something to make the rest disappear:

$$(X'X)^{-1}(X'X)\hat{\beta} = (X'X)^{-1}X'y$$
$$I\hat{\beta} = \hat{\beta} = (X'X)^{-1}X'y$$

- And now we've solved OLS's best estimates.

The Shape of Things to Come

Don't worry if the last few slides have seemed a little advanced. This is the kind of proof that takes a while to learn. The important thing is to understand how the building blocks we've covered: algebra, differentiation and optimization and vectors and matrices come together to solve an important problem. If we want to learn the relationships to be found in our data – the essence of data science and quantitative reasoning – these are the skills we'll need.

Thank you for your patience, and enjoy your semester!