MATH CAMP – SESSION 3

LINEAR ALGEBRA

Department of Economics

Duke University

Introduction

What is Linear Algebra?

- Linear algebra is about systems of linear equations, and can be with respect to any number of variables.
- These techniques are so powerful because we can apply them to many real-world problems, and they usually give us good solutions or approximations of solutions!
- In particular, linear algebra is focused on vectors.

Rules I

Definition (Scalar)

A *scalar* is a number (can be real or complex). We denote scalars with lowercase letters like a, b, c. Generally, scalars are denoted as $\mathbb{R}^{1\times 1}$ or simply $\mathbb{R}$.

Definition (Vector)

A *vector* is a collection of numbers in a single row or column. We denote vectors with boldface lowercase letters like $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$. The i^{th} element of a vector is usually denoted by $\mathbf{x}_i$.

Just like there are rules and operations on numbers, there are the same for vectors, such as addition and two types of multiplication (cross and dot product). We denote the set of "n by 1" vectors, that is the collections of n numbers, as $\mathbb{R}^{n\times 1}$, or more simply $\mathbb{R}^n$.

Rules II

Definition (Matrix)

A *matrix* is a collection of vectors. We denote matrices with uppercase letters like $\mathbf{A}$, $\mathbf{B}$, $\mathbf{P}$. The element in the i^{th} row and j^{th} column of a matrix is usually denoted a_{ij}.

Just like there are rules and operations on vectors and numbers, there are the same for matrices, such as addition, inversion, and multiplication. We denote the set of "n by m" matrices, that is the collections of m vectors with n numbers each, as $\mathbb{R}^{n \times m}$.

Why is it "algebra"?

Linear algebra is about applying various operations to numbers, vectors, and matrices. The algebra creates a set of rules for multiplication and addition, just like there are rules that you know for the real numbers. Adding matrices and vectors is easy! You just add element-wise.

Example 1.1:

$$\begin{bmatrix} -10 \\ 4 \\ 3 \\ -3 \end{bmatrix} + \begin{bmatrix} 8 \\ -4 \\ -1 \\ 5 \end{bmatrix} = \begin{bmatrix} -10+8 \\ 4+(-4) \\ 3+(-1) \\ -3+5 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 2 \\ 2 \end{bmatrix}.$$

Similarly, scalar multiplication is also easy, since it's element-wise.

Example 1.2:

$$3 \cdot \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \cdot 3 \\ 3 \cdot 2 \\ 3 \cdot 1 \end{bmatrix} = \begin{bmatrix} 9 \\ 6 \\ 3 \end{bmatrix}.$$

Linear Combinations

Definition (Linear Combination)

A *linear combination* of numbers, vectors, or matrices (or in general, objects) is an expression that only involves scalar multiplication and addition.

We would say "$3x + 4y$" is a linear combination of x and y, where x and y can be numbers, vectors, or matrices. This is because we are multiplying by scalars (3 and 4) and then adding them.

Example 1.3: Let $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$. A linear combination of these vectors with scalar coefficients a and b can be written as:

$$\mathbf{v} = a\mathbf{v}_1 + b\mathbf{v}_2 = a\begin{bmatrix} 1 \\ 2 \end{bmatrix} + b\begin{bmatrix} 3 \\ 4 \end{bmatrix}.$$

Vector Spaces

Definition (Vector Space)

Suppose $a, b \in \mathbb{R}$ are scalars and x, y, z are arbitrary objects (could be vectors). A *vector space* or *linear space* is a set of objects (does not have to be of vectors, despite the name) that is equipped with at least two operations: addition, $+$, and scalar multiplication, $\cdot$. It must also have the following "nice" properties:

1. Commutativity of $+$, i.e., $x + y = y + x$.
2. Associativity of $+$, i.e., $(x + y) + z = x + (y + z)$.
3. Associativity of $\cdot$, i.e., $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.
4. Neutral element of $+$, i.e., $x + 0 = x$
5. Neutral element of $\cdot$, i.e., $x \cdot 1 = x$.
6. Inverse of $+$, i.e., $\forall x, \exists y$ such that $x + y = 0$.
7. Distributivity of $\cdot$ over $+$, i.e., $a \cdot (x + y) = a \cdot x + a \cdot y$.
8. Distributivity of $+$ over $\cdot$, i.e., $(a + b) \cdot x = a \cdot x + b \cdot x$.

Definition (Subspace)

A *subspace* is a subset of a vector space such that the subset is also a vector space.

Examples of Vector Spaces and Subspaces

Example 2.1: Take the set of vectors with n elements, i.e., $\mathbb{R}^n$. This set is a vector space (show the above properties, using what we discussed about addition and scalar multiplication).

Additionally, consider the set of vectors with n even elements, which we could denote $\mathbb{R}^{n\times 1}_{\text{even}}$ is a subspace of $\mathbb{R}^{n\times 1}$.

Example 2.2: Consider the set of all polynomials of degree less than or equal to n, denoted by $\mathbb{P}_n$. This set forms a vector space, as it satisfies the required properties.

Furthermore, the set of all polynomials of degree less than or equal to m, where $m < n$, denoted by $\mathbb{P}_m$, is a subspace of $\mathbb{P}_n$.

Example 2.3: Consider the space $C([a, b])$ of all continuous functions defined on the interval $[a, b]$. This space forms a vector space under pointwise addition and scalar multiplication.

Additionally, the space of all continuous functions that vanish at a particular point (e.g., $f(a) = 0$) is a subspace of $C([a, b])$.

Collection of Vectors

Span

Definition (Span)

The *span* of a set of vectors is a collection of all linear combinations of those vectors. So, for a set S, $span(S)$ is the set of linear combinations of the elements of S.

Example 3.1: Suppose we have the following set of vectors:

$$S_1 = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}.$$

The span of this set of vectors is

$$span(S_1) = \left\{ a \begin{bmatrix} 1 \\ 1 \end{bmatrix} + b \begin{bmatrix} 1 \\ 0 \end{bmatrix} : \forall a, b \in \mathbb{R} \right\} = \left\{ \begin{bmatrix} a + b \\ a \end{bmatrix} : \forall a, b \in \mathbb{R} \right\} = \mathbb{R}^2.$$

Linear Independence I

Definition (Linearly Independent)

A set of vectors is *linearly independent* if there is no nontrivial linear combination of vectors that is 0. To be clear, suppose that the set of vectors is

$$S = \{x_1, x_2, \dots, x_k\} \subseteq \mathbb{R}^{n\times 1}.$$

A generic linear combination of these vectors would be

$$c_1x_1 + c_2x_2 + \cdots + c_kx_k = \sum_{i=1}^{k} c_ix_i.$$

If there is a choice of $\{c_1, c_2, \dots, c_k\}$ such that the above linear combination is 0, then S is linearly dependent. If there is no such choice of $\{c_1, c_2, \dots, c_k\}$, then S is linearly independent.

Linear Indendence II

Example 3.2: Take S_1 from before. This set is linearly independent. Take a general linear combination $\{c_1, c_2\}$, so we have

$$c_1 \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} + c_2 \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} c_1 + c_2 \\ c_1 \end{bmatrix}.$$

If this were to equal 0, then we would have

$$\begin{bmatrix} c_1 + c_2 \\ c_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \implies \begin{cases} c_1 + c_2 = 0 \\ c_1 = 0 \end{cases} \implies c_1 = 0 \implies c_2 = 0.$$

Now take S_2 from before. This set is linearly dependent. For example, the linear combination $\{1, 1, -1\}$ is nontrivial and would get you 0:

$$1 \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 1 \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + (-1) \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Basis and Dimension

Definition (Basis)

A set of vectors is a *basis* for a vector space if it is linearly independent and spans that vector space. Note: bases are **not** unique.

A set of vectors S is a basis for a vector space V if and only if every vector in V is a unique linear combination of the vectors in S. Geometric implication: every point in the space can be reached in exactly one way by the set of vectors.

Definition (Dimension)

The *dimension* of a vector space (assuming it is finite) is the number of vectors in its bases. That is, every basis of a vector space has the same number of vectors which we call the dimension or *dim*.

Definition (Standard Basis and Standard Vectors)

A *standard vector* or an *elementary vector* is a vector with exactly one element as 1 and the rest 0. Generally, we use the notation e_i to suggest the i^{th} element is 1 and the remaining elements are 0. The standard basis is the set of elementary vectors spanning a vector space.

Examples

Example 3.3: The standard basis for $\mathbb{R}^3$ is

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

You can prove this formally by applying the above definitions. Note: $dim(\mathbb{R}^n) = n$.

Example 3.4: Consider the space $M_{2\times 2}(\mathbb{R})$ of all 2×2 matrices. A possible basis for this vector space is the set of matrices

$$E_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad E_3 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad E_4 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Since there are 4 vectors in the basis, $dim(M_{2\times 2}(\mathbb{R})) = 4$.

Orthogonality I

Definition (Dot Product)

The *dot product* is the sum of the element-wise products of a pair of vectors which are the same size. It essentially describes the extent to which two vectors point in the same direction. It has a formula for two vectors $a, b \in \mathbb{R}^n$:

$$a \cdot b = \sum_{i=1}^{n} a_i b_i.$$

Definition (Norm and Normal)

The *norm* of a vector is the square root of the dot product of the vector with itself. It essentially represents the length of the vector. A vector is called *normal* if it has a norm of 1. We use the notation $||\cdot||$ for a vector $a \in \mathbb{R}^n$:

$$||a|| = \sqrt{a \cdot a} = \sqrt{\sum_{i=1}^{n} a_i}.$$

Orthogonality II

If you use the second equation to define the norm, you can also define the dot product as

$$a \cdot b = ||a||\, ||b||\ \cos(\theta)$$

where θ is the minimal geometric angle between a and b.

Definition (Orthogonal)

A set of vectors is *orthogonal* if every pair of vectors has dot product 0. A dot product of 0 means they don't point in the same direction at all|think of the x and y axis which means that orthogonality represents 90° angles.

Consider the set of vectors

$$S_3 = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \end{bmatrix} \right\}.$$

S_3 is not orthogonal because

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 1 \end{bmatrix} = 1 \cdot 2 + 0 \cdot 1 = 2 + 0 = 2 \neq 0.$$

Matrices

What is a "matrix"?

One way to see matrices is as a representation of a linear system. Suppose we have the system of equations

$$\begin{cases} 3x + 4y + 2z = 9 \\ x - y = 10 \\ -x + y + z = 3 \end{cases}.$$

This system can be represented as a matrix equation as

$$\begin{bmatrix} 3 & 4 & 2 \\ 1 & -1 & 0 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 9 \\ 10 \\ 3 \end{bmatrix}.$$

- We will learn about matrix multiplication shortly, but you can then multiply this back out to recover the system above!
- Another way to think of matrices is as a special function from one vector space to another, also known as a linear transformation.

What is a "matrix"? (contd.)

- In the above example, we are going from $\mathbb{R}^3$ to $\mathbb{R}^3$ since $\begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3$ and after "applying" the matrix, we end up with $\begin{bmatrix} 9 \\ 10 \\ 3 \end{bmatrix} \in \mathbb{R}^3$.
- But you can create matrices that go from any vector space to any other vector space!
- Finally, you can think of matrices geometrically: they stretch/contract and change the direction of vectors.
- Importantly here, a matrix is a collection of vectors, so it can have properties like span, orthogonality, be a basis, etc.

Matrix Multiplication I

- There is only one rule of matrix multiplication: the number of *columns* on the left matrix has to match the number of rows on the *right* matrix.
- Additionally, when you multiply and $m \times n$ matrix to a $n \times p$ matrix, then the output will be $m \times p$.
- The easiest way to do matrix multiplication is using dot products.
- Note: Matrix multiplication is not commutative.

Example I

Example 4.1: Suppose we want to find

$$\begin{bmatrix} 3 & 5 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

First, we make sure that the rule is followed: here it is followed since the left has 2 columns and the right has 2 rows which are the same. The first matrix is 2×2 and the second matrix is 2×1 so we know the output will be 2×1. To get the first element, we need to do the dot product of the first row on the left with the first (and only) column on the right; similarly, for the second element, we do the dot product of the second row on the left with the first (and only) column on the right. Thus we get

$$\begin{bmatrix} 3 & 5 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \cdot 2 + 5 \cdot 1 \\ 1 \cdot 2 + 2 \cdot 1 \end{bmatrix} = \begin{bmatrix} 11 \\ 4 \end{bmatrix}.$$

Example II

Example 4.2: Suppose we want to find

$$\begin{bmatrix} 3 & 5 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 1 & 5 \end{bmatrix}.$$

First, we make sure that the rule is followed: here it is followed since the left has 2 columns and the right has 2 rows which are the same. The first matrix is 2×2 and the second matrix is 2×2 so we know the output will be 2×2.

To get the element in the first row and first column, we need to do the dot product of the first row on the left with the first column on the right; similarly, for the element in the second row and first column, we do the dot product of the second row on the left with the first column on the right; we continue similarly for the second column. Thus, we get

$$\begin{bmatrix} 3 & 5 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 1 & 5 \end{bmatrix} = \begin{bmatrix} 3 \cdot 2 + 5 \cdot 1 & 3 \cdot 3 + 5 \cdot 5 \\ 1 \cdot 2 + 2 \cdot 1 & 1 \cdot 3 + 2 \cdot 5 \end{bmatrix} = \begin{bmatrix} 11 & 34 \\ 4 & 13 \end{bmatrix}.$$

The ones we focus on the most in economics are matrix-vector multiplication, especially in the context of trying to solve an equation of the form $Ax = b$ using least squares.

The Identity Matrix

Definition (Identity Matrix)

The *identity matrix* of dimension n, denoted I_n or 1_n, is an $n \times n$ matrix of the elementary vectors, written in order from e_1 on the left to e_n on the right.

Example 4.3: The identity matrix of dimension 4 is

$$I_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The identity matrix is special because you can multiply it times any matrix (as long as the multiplication is valid) and the multiplication will work.

Spaces of a Matrix

Matrix-vector multiplication is what we will focus on. You can assume $A \in \mathbb{R}^{m \times n}$. This section is focused on the solutions $x \in \mathbb{R}^n$ such that $Ax = 0$.

Definition (Rowspace)

The *rowspace* of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, denoted $R(\mathbf{A})$, is the set of all linear combinations of the rows of $\mathbf{A}$ – in other words, it is the span of the rows of $\mathbf{A}$. The rowspace is a subspace of $\mathbb{R}^m$.

Definition (Nullspace)

The *nullspace* or *kernel* of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, denoted $N(\mathbf{A})$, is the set of vectors $x \in \mathbb{R}^n$ such that $Ax = 0$. Note $0 \in N(\mathbf{A})$, always. The nullspace is a subspace of $\mathbb{R}^n$.

Example

Example 4.4: Take the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

The rowspace is

$$R(\mathbf{A}) = \left\{ a \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + b \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + c \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} : \forall a, b, c \in \mathbb{R} \right\} = \left\{ \begin{bmatrix} a + c \\ a \\ c \end{bmatrix} : \forall a, b, c \in \mathbb{R} \right\}$$

Notice that

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 - 1 + 0 \\ 0 + 0 + 0 \\ 1 - 1 + 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Thus $\begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix} \in N(A)$. Since this vector is in $N(A)$, any multiple of this vector is also in $N(A)$ (check!).

The Rank-Nullity Theorem

Rank-Nullity Theorem: This theorem says that $\dim(R(\mathbf{A})) + \dim(N(\mathbf{A})) = m$. It is really powerful because it connects the rowspace and the nullspace – there is always a geometric connection between the spaces of a matrix. It also tells you that if you have k vectors in your row space, then you must have $m - k$ vectors in your nullspace! This means you know exactly how many more vectors you need to find to make a full basis.

Definition (Transpose)

The *transpose* of a matrix A, denoted by A^T is taken by flipping its rows and columns. That is if a_{ij} is the element of A in the i^{th} row and j^{th} column, then a_{ji} is the element of A^T in the i^{th} row and j^{th} column.

Matrix Inversion I

Definition (Determinant)

The *determinant* of a square matrix A, denoted $\det(A)$, can be calculated as follows:

1. If the matrix is 1×1, then the determinant is just $\det(A) = a_{11}$.
2. If the matrix is $n \times n$, then you have to do a cofactor expansion on row i. The j^{th} cofactor of a row i is

$$c_{ij} = (-1)^{i+j} \det\left(\tilde{A}_{ij}\right)$$

where $\tilde{A}_{ij}$ is the matrix after removing the i^{th} row and j^{th} column (so $\tilde{A}_{ij}$ is an $(n-1) \times (n-1)$ matrix). Thus the cofactor expansion on row i is

$$\det(A) = \sum_{j=1}^{n} c_{ij} a_{ij}$$

Note: The only one you should memorize is the determinant of a 2×2 matrix which is $\det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = ad - bc$.

Matrix Inversion II

Definition (Invertible Matrix)

A square matrix $A \in \mathbb{R}^{n \times n}$ is *invertible* if it has a nonzero determinant. This also means that $\forall b \in \mathbb{R}^n, \exists! x \in \mathbb{R}^n$ such that $Ax = b$. It also means that there is an inverse called $A^{-1} \in \mathbb{R}^{n \times n}$ such that $AA^{-1} = A^{-1}A = I_n$.
Note: For an orthogonal matrix Q, $Q^T = Q^{-1}$, i.e., its transpose is its inverse.

Example 4.5: Consider the matrix A given by:

$$A = \begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix}$$

Since the determinant is non-zero (verify!), A is invertible. The inverse of A, denoted by A^{-1}, can be calculated as:

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} 4 & -1 \\ -3 & 2 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 4 & -1 \\ -3 & 2 \end{bmatrix} = \begin{bmatrix} \frac{4}{5} & -\frac{1}{5} \\ -\frac{3}{5} & \frac{2}{5} \end{bmatrix}$$

We can verify that:

$$AA^{-1} = \begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} \frac{4}{5} & -\frac{1}{5} \\ -\frac{3}{5} & \frac{2}{5} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2.$$

Eigenvalues and Eigenvectors I

Here, we are interested in directions such that if I apply the matrix, all it does is expand in that direction - so the matrix acts as a scalar.

Definition (Eigenvalue and Eigenvector)

An *eigenvalue-eigenvector* pair, or simply *eigenpair*, of a matrix $A \in \mathbb{R}^{n \times n}$ is a pair with a scalar λ (could be complex) and a vector $x \in \mathbb{R}^n$ such that $Ax = \lambda x$. This means A just changes the length of x λ but does not change the direction of x. These are really important for diagonalizing matrices, which makes them much easier to put into computers, much easier to solve systems, etc.

If $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of some matrix $A \in \mathbb{R}^{n \times n}$, then here are some cool properties:

- $\prod_{i=1}^n \lambda_i = \det(A)$
- $\sum_{i=1}^n \lambda_i = \operatorname{trace}(A) = \sum_{i=1}^n a_{ii}$

Application: Solving Linear Systems (OLS)

We want to solve the system $Ax = b$ or if we can't find a solution, then we want to get as close as possible. Here is the setup:

Suppose we have a set of inputs $x^1, x^2, \dots, x^m \in \mathbb{R}^n$ and we know their outputs to be $y^1, y^2, \dots, y^m \in \mathbb{R}$ (think of x being a bunch of factors to determine your GPA, and y being your actual GPA). Then suppose we set up a linear model

$$y = \beta_1 x_1 + \beta_2 x_2 + \dots \beta_k x_n + \epsilon = \beta^T x + \epsilon$$

where β is a vector of the β_i 's and x is a vector of the x_i 's. The idea is that we can use the points we know to set up an estimate of what β should be.

Ordinary Least Squares II

We can define the given x 's as the rows of a matrix:

$$X = \begin{bmatrix} x^{1T} \\ x^{2T} \\ \vdots \\ x^{mT} \end{bmatrix} \in \mathbb{R}^{m \times n}.$$

Similarly, we can define the output as

$$y = \begin{bmatrix} y^{1} \\ y^{2} \\ \vdots \\ y^{m} \end{bmatrix} \in \mathbb{R}^{m}.$$

Thus we want to solve the system $X\beta = y$ for β as the unknown. We can prove, using calculus (and matrix differentiation, which is complicated), that the best estimate, if you use least squares as you error measurement, is

$$\hat{\beta} = \left(X^T X\right)^{-1} X^T y.$$