\[
\begin{pmatrix}
-1 & 2 \\
2 & 2
\end{pmatrix}
\begin{pmatrix}
2 \\
2
\end{pmatrix}
= \ ?
\]

\[
A = \begin{pmatrix}
0 & 1 \\
1 & -1 \\
1 & 0
\end{pmatrix}
A^T = \ ?
\]

\[
A = \begin{bmatrix}
1 & 2 \\
3 & 4 \\
5 & 6
\end{bmatrix}
\text{ and } B = \begin{bmatrix}
7 & 10 \\
8 & 11 \\
9 & 12
\end{bmatrix}
\]

\[
A = \begin{bmatrix}
1 & 4 & 7 \\
2 & 5 & 8 \\
3 & 6 & 9
\end{bmatrix}
\text{ and } B = \begin{bmatrix}
1 & 4 \\
2 & 5 \\
3 & 6
\end{bmatrix}
\]

\[
\begin{pmatrix}
1 & 2 \\
-1 & 0
\end{pmatrix} + \begin{pmatrix}
1 & 0 \\
-1 & 1
\end{pmatrix}^T = \ ?
\]

\[
C = A - B = \ ?
\]

\[
C = A + B = \ ?
\]

\[
C = A \cdot B = ?
\]

\[
C = B \cdot A = ?
\]
Today:

- Data Representation for ML systems
- Review of Linear Algebra and Matrix Calculus
e.g. SUPERVISED LEARNING

• Find function to map **input** space \( X \) to **output** space \( Y \)

• **Generalisation**: learn function / hypothesis from **past data** in order to “explain”, “predict”, “model” or “control” **new** data examples
**Traditional Programming**

Data → Computer → Output

Program → Computer

---

**Machine Learning**

Data → Computer → Program / Model

Data → Output
An Operational Model of Machine Learning Software

Consists of input-output pairs

Reference Data → Learner → Model f

Tagged Data

Production Data → Execution Engine → Model f

Deployment
This Course: Before Deployment

Low-level sensing → Pre-processing → Feature Extract → Feature Select → Label Collection

Training: Optimization

\[ f : X \rightarrow Y \]

Testing: Inference, Prediction, Recognition

Evaluation

e.g. Data Cleaning
A Dataset

<table>
<thead>
<tr>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
<th>$Y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s_2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s_3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s_4$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s_5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s_6$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Data/points/instances/examples/samples/records**: [rows]
- **Features/attributes/dimensions/independent variables/covariates/predictors/regressors**: [columns, except the last]
- **Target/outcome/response/label/dependent variable**: special column to be predicted [last column]
Main Types of Columns

- **Continuous**: a real number, for example, weight

- **Discrete**: a symbol, like “Good” or “Bad”
**e.g. SUPERVISED Classification**

Training dataset consists of **input-output** pairs

- e.g. Here, target $Y$ is a **discrete** target variable
**e.g. SUPERVISED Regression**

Training dataset consists of **input-output** pairs

- **e.g.** Here, target $Y$ is a **continuous** target variable

$$f(x)$$
SUPERVISED LEARNING

Training dataset consists of input-output pairs

Evaluation

Measure Loss on pair \(\text{Error}(f(x_\text{?}), y_\text{?})\)
Machine Learning Variations in a Nutshell

ML grew out of work in AI

Optimize a performance criterion using example data or past experience,

Aiming to generalize to unseen data
Today:

- Data Representation for ML systems
- Review of Linear Algebra and Matrix Calculus
DEFINITIONS - SCALAR

◆ a scalar is a number
  – (denoted with regular type: 1 or 22)
DEFINITIONS - VECTOR

◆ **Vector**: a single row or column of numbers
  – denoted with **bold small letters**
  – row vector
    \[
    \mathbf{a} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \end{bmatrix}
    \]
  – column vector (default)
    \[
    \mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{bmatrix}
    \]
DEFINITIONS - VECTOR

- **Vector** in real space $\mathbb{R}^n$ is an ordered set of $n$ real numbers.
  - e.g. $\mathbf{v} = (1, 6, 3, 4)^T$ is in $\mathbb{R}^4$
  - A column vector:
  - $\mathbf{v}^T$ as a row vector:
    
    \[
    \begin{pmatrix}
    1 \\
    6 \\
    3 \\
    4
    \end{pmatrix}
    \]
    
    \[
    \begin{pmatrix}
    1 & 6 & 3 & 4
    \end{pmatrix}
    \]
DEFINITIONS - MATRIX

• m-by-n **matrix** in $\mathbb{R}^{mxn}$ with m rows and n columns, each entry filled with a (typically) real number:

• e.g. 3*3 matrix

\[
\begin{pmatrix}
1 & 2 & 8 \\
4 & 78 & 6 \\
9 & 3 & 2
\end{pmatrix}
\]

Square matrix
We normally write the entry of a matrix as

\[ A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \]

Denoted with a Capital letter

All matrices have an order (or dimension): that is, the number of rows * the number of columns.

So, \( A \) is 2 by 3 or \((2 \times 3)\).

A square matrix is a matrix that has the same number of rows and columns \((n \times n)\).
Special matrices

\[
\begin{pmatrix}
a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c
\end{pmatrix}
\]
diagonal

\[
\begin{pmatrix}
a & b & c \\ 0 & d & e \\ 0 & 0 & f
\end{pmatrix}
\]
upper-triangular

\[
\begin{pmatrix}
a & b & 0 & 0 \\ c & d & e & 0 \\ 0 & f & g & h \\ 0 & 0 & i & j
\end{pmatrix}
\]
tri-diagonal

\[
\begin{pmatrix}
a & 0 & 0 \\ b & c & 0 \\ d & e & f
\end{pmatrix}
\]
lower-triangular

\[
\begin{pmatrix}
1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1
\end{pmatrix}
\]
I (identity matrix)
Special matrices: Symmetric Matrices

\[ A = A^T \ (a_{ij} = a_{ji}) \]

e.g.:

\[
\begin{bmatrix}
4 & 5 & -3 \\
5 & 7 & 2 \\
-3 & 2 & 10
\end{bmatrix}
\]
Column or Row Views to Denote

- We denote the $j$th column of $A$ by $a_j$ or $A_{:,j}$:

\[
A = \begin{bmatrix}
    a_{11} & a_{12} & \cdots & a_{1n} \\
    a_{21} & a_{22} & \cdots & a_{2n} \\
    \vdots & \vdots & \ddots & \vdots \\
    a_{m1} & a_{m2} & \cdots & a_{mn}
\end{bmatrix}.
\]

- We denote the $i$th row of $A$ by $a_i^T$ or $A_{i,:}$:

\[
A = \begin{bmatrix}
    - & a_1^T & - \\
    - & a_2^T & - \\
    \vdots & \vdots & \ddots \\
    - & a_m^T & -
\end{bmatrix}.
\]

- Note that these definitions are ambiguous (for example, the $a_1$ and $a_1^T$ in the previous two definitions are not the same vector). Usually the meaning of the notation should be obvious from its use.
Review of MATRIX OPERATIONS

1) Transposition
2) Addition and Subtraction
3) Multiplication
4) Norm (of vector)
5) Matrix Inversion
6) Matrix Rank
7) Matrix calculus
(1) Transpose

**Transpose:** You can think of it as
- “flipping” the rows and columns

**Example:**

\[
\begin{pmatrix}
a \\
b \\
c \\
d
\end{pmatrix}
^T
= \begin{pmatrix}
a & b \\
\end{pmatrix}
\]

- \((A^T)^T = A\)
- \((AB)^T = B^T A^T\)
- \((A + B)^T = A^T + B^T\)
(2) Matrix Addition/Subtraction

• Matrix addition/subtraction
  – Matrices must be of same size.
  – Entry-wise operation across all entries
(2) Matrix Addition/Subtraction

An Example

• If we have

\[
A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 7 & 10 \\ 8 & 11 \\ 9 & 12 \end{bmatrix}
\]

then we can calculate \( C = A + B \) by

\[
C = A + B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} + \begin{bmatrix} 7 & 10 \\ 8 & 11 \\ 9 & 12 \end{bmatrix} = \begin{bmatrix} 8 & 12 \\ 11 & 15 \\ 14 & 18 \end{bmatrix}
\]
Similarly, if we have

\[
A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 7 & 10 \\ 8 & 11 \\ 9 & 12 \end{bmatrix}
\]

then we can calculate \( C = A - B \) by

\[
C = A - B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} - \begin{bmatrix} 7 & 10 \\ 8 & 11 \\ 9 & 12 \end{bmatrix} = \begin{bmatrix} -6 & -8 \\ -5 & -7 \\ -4 & -6 \end{bmatrix}
\]
OPERATION on MATRIX

1) Transposition
2) Addition and Subtraction
3) Multiplication
4) Norm (of vector)
5) Matrix Inversion
6) Matrix Rank
7) Matrix calculus
(3) Products of Matrices

• We write the multiplication of two matrices $A$ and $B$ as $AB$

• This is referred to either as
  • pre-multiplying $B$ by $A$
  or
  • post-multiplying $A$ by $B$

• So for matrix multiplication $AB$, $A$ is referred to as the *premultiplier* and $B$ is referred to as the *postmultiplier*
Products of Matrices

• If we have $A_{(3x3)}$ and $B_{(3x2)}$ then

$$AB = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \times \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ c_{31} & c_{32} \end{bmatrix} = C$$

where

$$
\begin{align*}
c_{11} &= a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} \\
c_{12} &= a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\
c_{21} &= a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} \\
c_{22} &= a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \\
c_{31} &= a_{31}b_{11} + a_{32}b_{21} + a_{33}b_{31} \\
c_{32} &= a_{31}b_{12} + a_{32}b_{22} + a_{33}b_{32}
\end{align*}
$$
Matrix Multiplication
An Example

• If we have

\[ \mathbf{A} = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \]

then

\[ \mathbf{AB} = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix} \times \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ c_{31} & c_{32} \end{bmatrix} = \begin{bmatrix} 30 & 66 \\ 36 & 81 \\ 42 & 96 \end{bmatrix} \]

where

\[ c_{11} = a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} = 1(1) + 4(2) + 7(3) = 30 \]

\[ c_{12} = a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} = 1(4) + 4(5) + 7(6) = 66 \]

\[ c_{21} = a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} = 2(1) + 5(2) + 8(3) = 36 \]

\[ c_{22} = a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} = 2(4) + 5(5) + 8(6) = 81 \]

\[ c_{31} = a_{31}b_{11} + a_{32}b_{21} + a_{33}b_{31} = 3(1) + 6(2) + 9(3) = 42 \]

\[ c_{32} = a_{31}b_{12} + a_{32}b_{22} + a_{33}b_{32} = 3(4) + 6(5) + 9(6) = 96 \]
Products of Matrices

\[
\begin{align*}
\text{m x n} & \quad \text{q x p} & \quad \text{m x p} \\
\begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{m1} & a_{m2} & \cdots & a_{mn}
\end{bmatrix} & \quad \begin{bmatrix}
b_{11} & b_{12} & \cdots & b_{1p} \\
b_{21} & b_{22} & \cdots & b_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
b_{q1} & b_{q2} & \cdots & b_{qp}
\end{bmatrix} & \quad \begin{bmatrix}
c_{11} & c_{12} & \cdots & c_{1p} \\
c_{21} & c_{22} & \cdots & c_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
c_{m1} & c_{m2} & \cdots & c_{mp}
\end{bmatrix}
\end{align*}
\]

**Condition:** \( n = q \)  \[ c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} \]

\( AB \neq BA \)
(3) Products of Matrices

• In order to multiply matrices, they must be conformable (the number of columns in the premultiplier must equal the number of rows in postmultiplier)

• Note that
  • an \( (m \times n) \times (n \times p) = (m \times p) \)
  • an \( (m \times n) \times (p \times n) \) = cannot be done
  • a \( (1 \times n) \times (n \times 1) \) = a scalar \( (1 \times 1) \)
Some Properties of Matrix Multiplication

• Note that

• Even if conformable, $AB$ does not necessarily equal $BA$ (i.e., matrix multiplication is not commutative)

• Matrix multiplication can be extended beyond two matrices

• matrix multiplication is associative, i.e., $A(BC) = (AB)C$
Some Properties of Matrix Multiplication

◆ Multiplication and transposition

$(AB)^T = B^T A^T$

◆ Multiplication with Identity Matrix

$AI = IA = A$, where $I = \begin{bmatrix}
1 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 1
\end{bmatrix}$
Special Uses for Matrix Multiplication

• **Products of Scalars & Matrices** ➔ Example, If we have

\[
A = \begin{bmatrix}
  1 & 2 \\
  3 & 4 \\
  5 & 6 \\
\end{bmatrix}
\]

and \( b = 3.5 \)

then we can calculate \( bA \) by

\[
bA = 3.5 \begin{bmatrix}
  1 & 2 \\
  3 & 4 \\
  5 & 6 \\
\end{bmatrix} = \begin{bmatrix}
  3.5 & 7.0 \\
  10.5 & 14.0 \\
  17.5 & 21.0 \\
\end{bmatrix}
\]

° Note that \( bA = Ab \) if \( b \) is a scalar
Special Uses for Matrix Multiplication

• **Dot (or Inner) Product** of two Vectors
  - Premultiplication of a column vector \( \mathbf{a} \) by conformable row vector \( \mathbf{b} \) yields a single value called the *dot product* or *inner product*
  - If
    \[
    \mathbf{b} = \begin{bmatrix} 5 \\ 2 \\ 8 \end{bmatrix} \quad \mathbf{a}^T = \begin{bmatrix} 3 & 4 & 6 \end{bmatrix}
    \]
    then their **inner product** gives us
    \[
    \mathbf{a}^T \mathbf{b} = \mathbf{a} \cdot \mathbf{b} = \begin{bmatrix} 3 & 4 & 6 \end{bmatrix} \begin{bmatrix} 5 \\ 2 \\ 8 \end{bmatrix} = 3(5) + 4(2) + 6(8) = 71 = \mathbf{b}^T \mathbf{a}
    \]
    which is the sum of products of elements in similar positions for the two vectors
Special Uses for Matrix Multiplication

• Outer Product of two Vectors
  • Postmultiplication of a column vector \( \mathbf{a} \) by conformable row vector \( \mathbf{b} \) yields a matrix containing the products of each pair of elements from the two matrices (called the outer product) - If

\[
\mathbf{a} = \begin{bmatrix} 3 \\ 4 \\ 6 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 5 \\ 2 \\ 8 \end{bmatrix}
\]

then \( \mathbf{a} \mathbf{b}^T \) gives us

\[
\mathbf{a} \mathbf{b}^T = \begin{bmatrix} 3 & 4 & 6 \\ 5 & 2 & 8 \end{bmatrix} = \begin{bmatrix} 15 & 6 & 24 \\ 20 & 8 & 32 \\ 30 & 12 & 48 \end{bmatrix}
\]
Special Uses for Matrix Multiplication

• Outer Product of two Vectors, e.g. a special case:

As an example of how the outer product can be useful, let \( \mathbf{1} \in \mathbb{R}^n \) denote an \( n \)-dimensional vector whose entries are all equal to 1. Furthermore, consider the matrix \( A \in \mathbb{R}^{m \times n} \) whose columns are all equal to some vector \( x \in \mathbb{R}^m \). Using outer products, we can represent \( A \) compactly as,

\[
A = \begin{bmatrix}
\vdots & \cdots & \vdots \\
x & x & \cdots & x \\
\vdots & \cdots & \vdots \\
x_m & x_m & \cdots & x_m \\
\end{bmatrix} = \begin{bmatrix}
x_1 & x_1 & \cdots & x_1 \\
x_2 & x_2 & \cdots & x_2 \\
\vdots & \vdots & \ddots & \vdots \\
x_m & x_m & \cdots & x_m \\
\end{bmatrix} = \begin{bmatrix}
x_1 \\
x_2 \\
\vdots \\
x_m \\
\end{bmatrix} \begin{bmatrix}
1 & 1 & \cdots & 1
\end{bmatrix} = x\mathbf{1}^T.
\]
Special Uses for Matrix Multiplication

• Matrix-Vector Products (I)

Given a matrix $A \in \mathbb{R}^{m \times n}$ and a vector $x \in \mathbb{R}^n$, their product is a vector $y = Ax \in \mathbb{R}^m$.

If we write $A$ by rows, then we can express $Ax$ as,

$$y = Ax = \begin{bmatrix}
  - & a_1^T & - \\
  - & a_2^T & - \\
  \vdots & \vdots & \vdots \\
  - & a_m^T & - \\
\end{bmatrix} x = \begin{bmatrix}
  a_1^T x \\
  a_2^T x \\
  \vdots \\
  a_m^T x \\
\end{bmatrix}.$$
Special Uses for Matrix Multiplication

• **Matrix-Vector Products (II)**

Alternatively, let’s write $A$ in column form. In this case we see that,

$$
y = Ax = \begin{bmatrix} \vdots \\
a_1 \\
a_2 \\
\vdots \\
a_n \\
\end{bmatrix} \begin{bmatrix} x_1 \\
x_2 \\
\vdots \\
x_n \\
\end{bmatrix} = \begin{bmatrix} a_1 \\
a_2 \\
\vdots \\
a_n \\
\end{bmatrix} x_1 + \begin{bmatrix} a_2 \\
a_3 \\
\vdots \\
a_n \\
\end{bmatrix} x_2 + \cdots + \begin{bmatrix} a_n \\
\end{bmatrix} x_n.
$$

In other words, $y$ is a **linear combination** of the columns of $A$, where the coefficients of the linear combination are given by the entries of $x$. 
to multiply on the left by a row vector. This is written, $y^T = x^T A$ for $A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^m$, and $y \in \mathbb{R}^n$.

$$y^T = x^T A = x^T \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} = \begin{bmatrix} x^T a_1 & x^T a_2 & \cdots & x^T a_n \end{bmatrix}$$

which demonstrates that the $i$th entry of $y^T$ is equal to the inner product of $x$ and the $i$th column of $A$. 

4/5/18

Dr. Yanjun Qi / 42
Special Uses for Matrix Multiplication

• **Matrix-Vector Products (IV)**

\[
y^T = x^T A
\]

\[
= \begin{bmatrix}
x_1 & x_2 & \cdots & x_n
\end{bmatrix}
\begin{bmatrix}
\vdots \\
- a_1^T & - \\
- a_2^T & - \\
\vdots \\
- a_m^T & - \\
\end{bmatrix}
\]

\[
= x_1 \begin{bmatrix}
- a_1^T & - 
\end{bmatrix} + x_2 \begin{bmatrix}
- a_2^T & - 
\end{bmatrix} + \cdots + x_n \begin{bmatrix}
- a_n^T & - 
\end{bmatrix}
\]

so we see that \( y^T \) is a linear combination of the *rows* of \( A \), where the coefficients for the linear combination are given by the entries of \( x \).
MATRIX OPERATIONS

1) Transposition
2) Addition and Subtraction
3) Multiplication
4) Norm (of vector)
5) Matrix Inversion
6) Matrix Rank
7) Matrix calculus
(4) Vector norms

A norm of a vector $\|x\|$ is informally a measure of the “length” of the vector.

$$\|x\|_p = \left( \sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

- Common norms: $L_1$, $L_2$ (Euclidean)

$$\|x\|_1 = \sum_{i=1}^{n} |x_i| \quad \|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{x^\top x}$$

- $L_{\infty}$

$$\|x\|_{\infty} = \max_i |x_i|$$
Vector Norm (L2, when p=2)

\[ \left\| \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\|_2 = \sqrt{1^2 + 2^2} = \sqrt{5} \]
Special Uses for Matrix Multiplication

• **Sum the Squared Elements of a Vector**
  - Premultiply a column vector $\mathbf{a}$ by its transpose
    - If
      $$\mathbf{a} = \begin{bmatrix} 5 \\ 2 \\ 8 \end{bmatrix}$$
      then premultiplication by a row vector $\mathbf{a}^T$
      $$\mathbf{a}^T = \begin{bmatrix} 5 & 2 & 8 \end{bmatrix}$$
      will yield the sum of the squared values of elements for $\mathbf{a}$, i.e.
      $$\mathbf{a}^T \mathbf{a} = \begin{bmatrix} 5 & 2 & 8 \end{bmatrix} \begin{bmatrix} 5 \\ 2 \\ 8 \end{bmatrix} = 5^2 + 2^2 + 8^2 = 93$$
Vector Norms (e.g.,)

Drawing shows unit sphere in two dimensions for each norm

\((-1.6, 1.2)\)

Norms have following values for vector shown

\[
\|x\|_1 = 2.8 \quad \|x\|_2 = 2.0 \quad \|x\|_\infty = 1.6
\]

In general, for any vector \(x\) in \(\mathbb{R}^n\), \(\|x\|_1 \geq \|x\|_2 \geq \|x\|_\infty\)
More General : Norm

- A norm is any function $g()$ that maps vectors to real numbers that satisfies the following conditions:
  
  - **Non-negativity**: for all $x \in \mathbb{R}^D$, $g(x) \geq 0$
  - **Strictly positive**: for all $x$, $g(x) = 0$ implies that $x = 0$
  - **Homogeneity**: for all $x$ and $a$, $g(ax) = |a| g(x)$, where $|a|$ is the absolute value.
  - **Triangle inequality**: for all $x, y$, $g(x + y) \leq g(x) + g(y)$
Orthogonal & Orthonormal

Inner Product defined between column vector $\mathbf{x}$ and $\mathbf{y}$, as

$$x^T y \in \mathbb{R} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ x_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i = \mathbf{x} \cdot \mathbf{y}$$

If $\mathbf{u} \cdot \mathbf{v} = 0$, $||\mathbf{u}||_2 \neq 0$, $||\mathbf{v}||_2 \neq 0$

$\rightarrow$ $\mathbf{u}$ and $\mathbf{v}$ are orthogonal

If $\mathbf{u} \cdot \mathbf{v} = 0$, $||\mathbf{u}||_2 = 1$, $||\mathbf{v}||_2 = 1$

$\rightarrow$ $\mathbf{u}$ and $\mathbf{v}$ are orthonormal
Orthogonal matrices

- **Notation:**

\[
A = \begin{bmatrix}
  a_{11} & a_{12} & \cdots & a_{1n} \\
  a_{21} & a_{22} & \cdots & a_{2n} \\
  \vdots & \vdots & \ddots & \vdots \\
  a_{m1} & a_{m2} & \cdots & a_{mn}
\end{bmatrix},
\]

\[
\begin{align*}
  u_1^T &= [a_{11} \ a_{12} \cdots a_{1n}] \\
  u_2^T &= [a_{21} \ a_{22} \cdots a_{2n}] \\
  \vdots & \quad \vdots & \quad \ddots & \quad \vdots \\
  u_m^T &= [a_{m1} \ a_{m2} \cdots a_{mn}]
\end{align*}
\]

\[
A = \begin{bmatrix}
  u_1^T \\
  u_2^T \\
  \vdots \\
  u_m^T
\end{bmatrix}
\]

- **A is orthogonal if:**

1. \(u_k \cdot u_k = 1\) or \(||u_k|| = 1\), for every \(k\)

2. \(u_j \cdot u_k = 0\), for every \(j \neq k\) (\(u_j\) is perpendicular to \(u_k\))

**Example:**

\[
\begin{bmatrix}
  \cos(\theta) & -\sin(\theta) \\
  \sin(\theta) & \cos(\theta)
\end{bmatrix}
\]
Orthogonal matrices

• If square $A$ is orthogonal, it is easy to find its inverse:

\[ AA^T = A^T A = I \quad \text{(i.e., } A^{-1} = A^T)\]

Property: $\|Av\| = \|v\|$ (does not change the magnitude of $v$)
Matrix Norm

• **Definition**: Given a vector norm \( ||x|| \), the **matrix norm** defined by the vector norm is given by:

\[
||A|| = \max_{x \neq 0} \frac{||Ax||}{||x||}
\]

• What does a matrix norm represent?
• It represents the maximum “stretching” that A does to a vector \( x \rightarrow (Ax) \).
Matrix 1- Norm

**Theorem A**: The matrix norm corresponding to 1-norm is maximum absolute column sum:

\[ \| A \|_1 = \max_j \sum_{i=1}^n |a_{ij}| \]

**Proof**: From previous slide, we can have \( \| A \|_1 = \max_{\| x \| = 1} \| Ax \|_1 \)

Also, \( Ax = x_1 A_1 + x_2 A_2 + \cdots + x_n A_n = \sum_{j=1}^n x_j A_j \)

where \( A_j \) is the j-th column of \( A \).
MATRIX OPERATIONS

1) Transposition
2) Addition and Subtraction
3) Multiplication
4) Norm (of vector)
5) Matrix Inversion
6) Matrix Rank
7) Matrix calculus
(5) Inverse of a Matrix

• The inverse of a matrix \( A \) is commonly denoted by \( A^{-1} \) or \( \text{inv } A \).

• The inverse of an \( n \times n \) matrix \( A \) is the matrix \( A^{-1} \) such that \( AA^{-1} = I = A^{-1}A \).

• The matrix inverse is analogous to a scalar reciprocal.

• A matrix which has an inverse is called 

  \textit{nonsingular}
(5) Inverse of a Matrix

• For some \( n \times n \) matrix \( A \), an inverse matrix \( A^{-1} \) may not exist.

• A matrix which does not have an inverse is singular.

• An inverse of \( n \times n \) matrix \( A \) exists iff \( |A| \) not 0.
THE DETERMINANT OF A MATRIX

◆ The determinant of a matrix $A$ is denoted by $|A|$ (or $\det(A)$ or $\det A$).

◆ Determinants exist only for square matrices.

◆ E.g. If $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$

\[ |A| = a_{11}a_{22} - a_{12}a_{21} \]
THE DETERMINANT OF A MATRIX

2 x 2

\[ A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \text{det}(A) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12} \]

3 x 3

\[
\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}
\]

n x n

\[
\text{det}(A) = \sum_{j=1}^{m} (-1)^{j+k} a_{jk} \text{det}(A_{jk}), \text{ for any } k: 1 \leq k \leq m
\]
THE DETERMINANT OF A MATRIX

\[ \text{det}(AB) = \text{det}(A)\text{det}(B) \]

\[ \text{det}(A + B) \neq \text{det}(A) + \text{det}(B) \]

diagonal matrix: If \( A = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix} \), then \( \text{det}(A) = \prod_{i=1}^{n} a_{ii} \)
HOW TO FIND INVERSE MATRIXES?

An example,

\[ A \begin{bmatrix} a & b \\ c & d \end{bmatrix} \]

and \(|A|\) not 0

\[ A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \]
Matrix Inverse

• The inverse $A^{-1}$ of a matrix $A$ has the property:

\[ AA^{-1} = A^{-1}A = I \]

• $A^{-1}$ exists if only if $\det(A) \neq 0$

• Terminology
  
  – **Singular matrix:** $A^{-1}$ does not exist
  
  – **Ill-conditioned matrix:** $A$ is close to being singular
PROPERTIES OF INVERSE MATRICES

◆ \((AB)^{-1} = B^{-1}A^{-1}\)

◆ \((A^T)^{-1} = (A^{-1})^T\)

◆ \((A^{-1})^{-1} = A\)
Inverse of special matrix

• For diagonal matrices

\[ D^{-1} = \text{diag}\{d_1^{-1}, \ldots, d_n^{-1}\} \]

• For orthogonal matrices

\[ A^{-1} = A^\top \]

– a square matrix with real entries whose columns and rows are orthogonal unit vectors (i.e., orthonormal vectors)
Pseudo-inverse

• The pseudo-inverse \( A^+ \) of a matrix A (could be non-square, e.g., m x n) is given by:

\[
A^+ = (A^T A)^{-1} A^T
\]

• It can be shown that:

\[
A^+ A = I \quad \text{(provided that } (A^T A)^{-1} \text{ exists)}
\]
MATRIX OPERATIONS

1) Transposition
2) Addition and Subtraction
3) Multiplication
4) Norm (of vector)
5) Matrix Inversion
6) Matrix Rank
7) Matrix calculus
(6) Rank: Linear independence

- A set of vectors is **linearly independent** if none of them can be written as a linear combination of the others.

\[
\begin{align*}
x_1 &= \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} & x_2 &= \begin{bmatrix} 4 \\ 1 \\ 5 \end{bmatrix} & x_3 &= \begin{bmatrix} 2 \\ -3 \\ -1 \end{bmatrix}
\end{align*}
\]

\[x_3 = -2 \, x_1 + x_2\]

\(\Rightarrow\) **NOT linearly independent**
(6) **Rank: Linear independence**

- **Alternative definition:** Vectors \(v_1, \ldots, v_k\) are linearly independent if \(c_1 v_1 + \ldots + c_k v_k = 0\) implies \(c_1 = \ldots = c_k = 0\)

\[
\begin{pmatrix}
| & | & | \\
v_1 & v_2 & v_3 \\
| & | & | \\
c_1 & c_2 & c_3
\end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}
\]

e.g.

\[
\begin{pmatrix} 1 & 0 \\ 2 & 3 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}
\]

\((u,v) = (0,0)\), i.e. the columns are linearly independent.
(6) Rank of a Matrix

- rank(A) (the rank of a m-by-n matrix A) is
  - The maximal number of linearly independent columns
  - The maximal number of linearly independent rows

- If A is n by m, then
  - $\text{rank}(A) \leq \min(m,n)$
  - If $n = \text{rank}(A)$, then A has full row rank
  - If $m = \text{rank}(A)$, then A has full column rank
(6) Rank of a Matrix

• Equal to the dimension of the largest square sub-matrix of $A$ that has a non-zero determinant.

Example: \[
\begin{bmatrix}
4 & 5 & 2 & 14 \\
3 & 9 & 6 & 21 \\
8 & 10 & 7 & 28 \\
1 & 2 & 9 & 5
\end{bmatrix}
\]

has rank 3

\[\text{det}(A) = 0, \text{ but } \text{det}\left(\begin{bmatrix}
4 & 5 & 2 \\
3 & 9 & 6 \\
8 & 10 & 7
\end{bmatrix}\right) = 63 \neq 0\]
(6) Rank and singular matrices

If $A$ is $n \times n$, $\text{rank}(A) = n$ iff $A$ is nonsingular (i.e., invertible).

If $A$ is $n \times n$, $\text{rank}(A) = n$ iff $\det(A) \neq 0$ (full rank).

If $A$ is $n \times n$, $\text{rank}(A) < n$ iff $A$ is singular

We can use row reduction to calculating Rank of a matrix
From Wiki

The following complexity figures assume that arithmetic with individual elements has complexity $O(1)$, as is the case with fixed-precision operations on a finite field.

<table>
<thead>
<tr>
<th>Operation</th>
<th>Input</th>
<th>Output</th>
<th>Algorithm</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matrix multiplication</td>
<td>Two $n \times n$ matrices</td>
<td>One $n \times n$ matrix</td>
<td>Schoolbook matrix multiplication</td>
<td>$O(n^3)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Strassen algorithm</td>
<td>$O(n^{2.807})$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Coppersmith–Winograd algorithm</td>
<td>$O(n^{2.376})$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Optimized CW-like algorithms$^{[14][15][16]}$</td>
<td>$O(n^{2.373})$</td>
</tr>
<tr>
<td>Matrix multiplication</td>
<td>One $n \times m$ matrix &amp; one $m \times p$ matrix</td>
<td>One $m \times p$ matrix</td>
<td>Schoolbook matrix multiplication</td>
<td>$O(nmp)$</td>
</tr>
<tr>
<td>Matrix inversion$^*$</td>
<td>One $n \times n$ matrix</td>
<td>One $n \times n$ matrix</td>
<td>Gauss–Jordan elimination</td>
<td>$O(n^3)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Strassen algorithm</td>
<td>$O(n^{2.807})$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Coppersmith–Winograd algorithm</td>
<td>$O(n^{2.376})$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Optimized CW-like algorithms$^{[14][15][16]}$</td>
<td>$O(n^{2.373})$</td>
</tr>
<tr>
<td>Singular value decomposition</td>
<td>One $m \times n$ matrix</td>
<td>One $m \times m$ matrix, one $m \times n$ matrix, &amp; one $n \times n$ matrix</td>
<td>$O(mm^2)$ ($m \leq n$)</td>
<td></td>
</tr>
<tr>
<td>Determinant</td>
<td>One $n \times n$ matrix</td>
<td>One number</td>
<td>Laplace expansion</td>
<td>$O(n!)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Division-free algorithm$^{[17]}$</td>
<td>$O(n^4)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LU decomposition</td>
<td>$O(n^3)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bareiss algorithm</td>
<td>$O(n^3)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Fast matrix multiplication$^{[18]}$</td>
<td>$O(n^{2.373})$</td>
</tr>
<tr>
<td>Back substitution</td>
<td>Triangular matrix</td>
<td>$n$ solutions</td>
<td>Back substitution$^{[19]}$</td>
<td>$O(n^2)$</td>
</tr>
</tbody>
</table>
MATRIX OPERATIONS

1) Transposition
2) Addition and Subtraction
3) Multiplication
4) Norm (of vector)
5) Matrix Inversion
6) Matrix Rank
7) Matrix calculus
Review: Derivative of a Function

\[
\lim_{h \to 0} \frac{f(a+h) - f(a)}{h}
\]

is called the derivative of \( f \) at \( a \).

We write:

\[
f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}
\]

“The derivative of \( f \) with respect to \( x \) is ...”

There are many ways to write the derivative of \( y = f(x) \)

\[\Rightarrow\text{ e.g. define the slope of the curve } y=f(x) \text{ at the point } x\]
Review: Derivative of a Quadratic Function

\[ y = x^2 - 3 \]

\[ y' = \lim_{h \to 0} \frac{(x + h)^2 - 3 - (x^2 - 3)}{h} \]

\[ y' = \lim_{h \to 0} \frac{x^2 + 2xh + h^2 - x^2}{h} \]

\[ y' = \lim_{h \to 0} 2x + h \]

\[ y' = 2x \]
# Single Var-Func to Multivariate

<table>
<thead>
<tr>
<th>Single Var-Function</th>
<th>Multivariate Calculus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Derivative</td>
<td>Partial Derivative</td>
</tr>
<tr>
<td>Second-order</td>
<td>Gradient</td>
</tr>
<tr>
<td>derivative</td>
<td>Directional Partial</td>
</tr>
<tr>
<td></td>
<td>Derivative</td>
</tr>
<tr>
<td></td>
<td>Vector Field</td>
</tr>
<tr>
<td></td>
<td>Contour map of a</td>
</tr>
<tr>
<td></td>
<td>function</td>
</tr>
<tr>
<td></td>
<td>Surface map of a</td>
</tr>
<tr>
<td></td>
<td>function</td>
</tr>
<tr>
<td></td>
<td>Hessian matrix</td>
</tr>
<tr>
<td></td>
<td>Jacobian matrix (vector in / vector out)</td>
</tr>
</tbody>
</table>
Some important rules for taking (partial) derivatives

- **Scalar multiplication**: $\partial_x[af(x)] = a[\partial_x f(x)]$
- **Polynomials**: $\partial_x[x^k] = kx^{k-1}$
- **Function addition**: $\partial_x[f(x) + g(x)] = [\partial_x f(x)] + [\partial_x g(x)]$
- **Function multiplication**: $\partial_x[f(x)g(x)] = f(x)[\partial_x g(x)] + [\partial_x f(x)]g(x)$
- **Function division**: $\partial_x \left[ \frac{f(x)}{g(x)} \right] = \frac{[\partial_x f(x)]g(x) - f(x)[\partial_x g(x)]}{[g(x)]^2}$
- **Function composition**: $\partial_x[f(g(x))] = [\partial_x g(x)][\partial_x f](g(x))$
- **Exponentiation**: $\partial_x[e^x] = e^x$ and $\partial_x[a^x] = \log(a)e^x$
- **Logarithms**: $\partial_x[\log x] = \frac{1}{x}$
Suppose that \( f : \mathbb{R}^{m \times n} \rightarrow \mathbb{R} \) is a function that takes as input a matrix \( A \) of size \( m \times n \) and returns a real value. Then the **gradient** of \( f \) (with respect to \( A \in \mathbb{R}^{m \times n} \)) is the matrix of partial derivatives:

\[
\nabla_A f(A) \in \mathbb{R}^{m \times n} = \begin{bmatrix}
\frac{\partial f(A)}{\partial A_{11}} & \frac{\partial f(A)}{\partial A_{12}} & \cdots & \frac{\partial f(A)}{\partial A_{1n}} \\
\frac{\partial f(A)}{\partial A_{21}} & \frac{\partial f(A)}{\partial A_{22}} & \cdots & \frac{\partial f(A)}{\partial A_{2n}} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial f(A)}{\partial A_{m1}} & \frac{\partial f(A)}{\partial A_{m2}} & \cdots & \frac{\partial f(A)}{\partial A_{mn}}
\end{bmatrix}
\]

In principle, gradients are a natural extension of partial derivatives to functions of multiple variables.
Review: Definitions of gradient (Matrix_calculus / Scalar-by-vector)

- Size of gradient is always the same as the size of

\[ \nabla_x f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix} \in \mathbb{R}^n \text{ if } x \in \mathbb{R}^n \]
For Examples

\[
\frac{\partial x^T a}{\partial x} = \frac{\partial a^T x}{\partial x} = a
\]

\[
\frac{\partial a^T X b}{\partial X} = ab^T
\]

\[
\frac{\partial a^T X^T b}{\partial X} = ba^T
\]

\[
\frac{\partial a^T X a}{\partial X} = \frac{\partial a^T X^T a}{\partial X} = aa^T
\]

\[
\frac{\partial x^T B x}{\partial x} = (B + B^T)x
\]
Exercise: a simple example

\[
f(w) = w^T a = \begin{bmatrix} w_1, w_2, w_3 \end{bmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = w_1 + 2w_2 + 3w_3
\]

\[
\frac{\partial f}{\partial w_1} = 1 \\
\frac{\partial f}{\partial w_2} = 2 \\
\frac{\partial f}{\partial w_3} = 3
\]

\[
\frac{\partial f}{\partial w} = \frac{\partial w^T a}{\partial w} = a = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}
\]

"Denominator layout"
Even more general Matrix Calculus: Types of Matrix Derivatives

<table>
<thead>
<tr>
<th></th>
<th>Scalar</th>
<th>Vector</th>
<th>Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scalar</td>
<td>$\frac{df}{dx}$</td>
<td>$\frac{dF}{dx} = \left[ \frac{\partial F_i}{\partial x} \right]$</td>
<td>$\frac{dF}{dx} = \left[ \frac{\partial F_{ij}}{\partial x} \right]$</td>
</tr>
<tr>
<td>Vector</td>
<td>$\frac{df}{dX} = \left[ \frac{df}{dX_i} \right]$</td>
<td>$\frac{dF}{dX} = \left[ \frac{\partial F_i}{\partial X_j} \right]$</td>
<td></td>
</tr>
<tr>
<td>Matrix</td>
<td>$\frac{df}{dX} = \left[ \frac{df}{dX_{ij}} \right]$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Adapted from Thomas Minka. Old and New Matrix Algebra Useful for Statistics
Review: Hessian Matrix / n=2 case

Singlevariate $\rightarrow$ multivariate

- 1\textsuperscript{st} derivative to gradient,
  
  $$g = \nabla f = \left( \begin{array}{c} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial z} \end{array} \right)$$

- 2\textsuperscript{nd} derivative to Hessian

  $$H = \begin{pmatrix}
  \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial z} \\ 
  \frac{\partial^2 f}{\partial x \partial z} & \frac{\partial^2 f}{\partial z^2}
  \end{pmatrix}$$
Review: Hessian Matrix

Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a function that takes a vector in $\mathbb{R}^n$ and returns a real number. Then the Hessian matrix with respect to $x$, written $\nabla_x^2 f(x)$ or simply as $H$ is the $n \times n$ matrix of partial derivatives,

$$\nabla_x^2 f(x) \in \mathbb{R}^{n \times n} = \begin{bmatrix}
\frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\
\frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2}
\end{bmatrix}.$$
Today Recap

- Data Representation
- Linear Algebra and Matrix Calculus Review
  1) Transposition
  2) Addition and Subtraction
  3) Multiplication
  4) Norm (of vector)
  5) Matrix Inversion
  6) Matrix Rank
  7) Matrix calculus
Notation

• Inputs
  – $X$, $X_j$ (jth element of vector $X$) : random variables written in capital letter
  – $p$ #inputs, $N$ #observations
  – $X$ : matrix written in bold capital
  – Vectors are assumed to be column vectors
  – Discrete inputs often described by characteristic vector (dummy variables)

• Outputs
  – quantitative $Y$
  – qualitative $C$ (for categorical)

• Observed variables written in lower case
  – The i-th observed value of $X$ is $x_i$ and can be a scalar or a vector
• The following topics are covered by handout, but not by this slide (some will be covered ...)
  – Trace()
  – Eigenvalue / Eigenvectors
  – Positive definite matrix, Gram matrix
  – Quadratic form
  – Projection (vector on a plane, or on a vector)
Best Place to Review: Khan Academy

**Vectors and spaces**
0 of 45 complete
Let's get our feet wet by thinking in terms of vectors and spaces.

**Matrix transformations**
0 of 58 complete
Understanding how we can map one set of vectors to another set. Matrices used to define linear transformations.

**Alternate coordinate systems (bases)**
0 of 39 complete
We explore creating and moving between various coordinate systems.

**Linear algebra**
- Vectors
- Linear combinations and spans
- Linear dependence and independence
- Subspaces and the basis for a subspace
- Vector dot and cross products
- Matrices for solving systems by elimination
- Null space and column space
- Functions and linear transformations
- Linear transformation examples
- Transformations and matrix multiplication
- Inverse functions and transformations
- Finding inverses and determinants
- More determinant depth
- Transpose of a matrix
- Orthogonal complements
- Orthogonal projections
- Change of basis
- Orthonormal bases and the Gram-Schmidt process
- Eigen-everything
Thinking about multivariable functions
2 of 22 complete
The only thing separating multivariable calculus from ordinary calculus is this newfangled word "multivariable". It means we will deal with functions whose inputs or outputs live in two or more dimensions. Here we lay the foundations for thinking about and visualizing multivariable functions.

Derivatives of multivariable functions
6 of 72 complete
What does it mean to take the derivative of a function whose input lives in multiple dimensions? What about when its output is a vector? Here we go over many different ways to extend the idea of a derivative to higher dimensions, including partial derivatives, directional derivatives, the gradient, vector derivatives, divergence, curl, etc.

Applications of multivariable derivatives
1 of 37 complete
The tools of partial derivatives, the gradient, etc. can be used to optimize and approximate multivariable functions. These are very useful in practice, and to a large extent this is why people study multivariable calculus.
References

- Prof. James J. Cochran’s tutorial slides “Matrix Algebra Primer II”
- [http://www.cs.cmu.edu/~aarti/Class/10701/recitation/LinearAlgebra_Matlab_Review.ppt](http://www.cs.cmu.edu/~aarti/Class/10701/recitation/LinearAlgebra_Matlab_Review.ppt)
- Prof. Alexander Gray’s slides
- Prof. George Bebis’ slides
- Prof. Hal Daumé III’ notes
- Khan Academy