COMP 5630/6630:Machine Learning

Lecture 2: Linear Algebra

Slides are adapted from Drs. Aakur , Farhana, Andrew Ng and cs229 course at Stanford

Why Linear Algebra

- ML
  - Learn a “mapping” from input to output f: X → Y
    - Fitting a function, f, from input to output using the dataset
  - Learning objective: minimize the loss between “learned mapping” and “true mapping”

- Linear Algebra
  - Mathematical foundation to representing data as well as computations in ML

Linear Algebra in ML

- Data Representation: Matrices and vectors
- Preprocessing
- Dimensionality Reduction
- Feature Engineering
- Designing ML Algorithm
- Model Interpretation

Linear Algebra Topics

Basic Concepts and Notations	Scalars, Matrices, Vectors, and Tensors Identity and diagonal matrix
Operations and Properties	Transpose Inverse Matrix and Vector Multiplication Linear Transform and Span Norms Special kinds of matrices and vectors Eigen decomposition

Scalar

- Scalar
  - A single number: real-valued or integers
  - Represented in lower-case italic x
    - i.e. slope of a line (real-valued)

Matrix

- A 2-D array of numbers arranged in order
- Denoted by bold typeface **A**

- Dimension of a matrix: number of rows x number of columns

  $\mathbf{A} \in \mathbb{R}^{m \times n}$**;** m rows and n columns

- Elements indicated by name in italic but not bold

- $\mathbf{A}_{ij}$ represents the element in the i^{th} row and j^{th} column

Matrix Example

- $\mathbf{A} = \begin{bmatrix} 12 & 34 \\ -123 & 15 \\ 145 & 10 \end{bmatrix}$

- a_{11} =12 a_{12}= 34
- $a_{21} = -123$ a_{22} = 15
- $a_{31} = 145$ a_{32} = 10

Vector

- A 1-D array of numbers
- Written in lower-case bold such as ***x***
- A vector is an n x 1 dimensional matrix

$$\mathbf{x} = \begin{bmatrix} 25 \\ 145 \\ -30 \end{bmatrix}$$

n= 3

$\mathbf{x}_i$ = i_{th} element

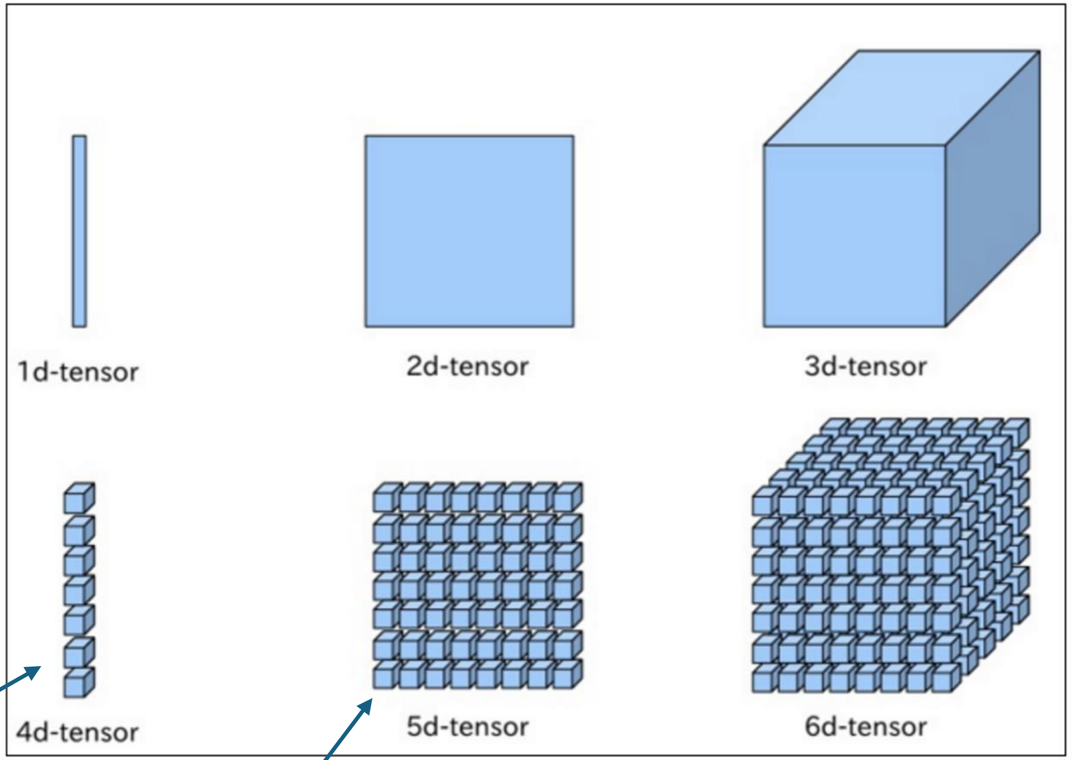
$\mathbf{x}_1$ = 25

$\mathbf{x}_2$ = 145

Tensors

- Array of numbers with more than two dimensions

- Denoted by bold typeface **A**. Elements given by $\mathbf{A}_{ijk}$ for 3-d tensor.



3d-tensors arranged in one grid= 4d

3d-tensors arranged in 2-grids= 5d

Identity Matrix

- A square matrix with ones on the diagonal and zeros everywhere else
- Denoted by $I \in \mathbb{R}^{n \times n}$

- Properties

1. $\boldsymbol{I}_{ij} = \begin{cases} 1, i = j \\ 0 \; i \neq j \end{cases}$

2. **AI** = **A** = **IA,** where $\mathbf{A} \in \mathbb{R}^{n \times n}$

Diagonal Matrix

- A matrix with all non-diagonal elements are 0
- Denoted by D = diag(d_1, d_2,…, d_n), where
- $\mathbf{D}_{ij}$ = $\begin{cases} d_i, i = j \\ 0\ i \neq j \end{cases}$
- Identity matrix, I = ?

Diagonal Matrix

- A matrix with all non-diagonal elements are 0

- Denoted by D = diag(d_1, d_2,…, d_n), where

- $\mathbf{D}_{ij}$ = $\begin{cases} d_i, i = j \\ 0\ i \neq j \end{cases}$

- Identity matrix, I = diag(1, 1,…, 1)

Matrix/Tensor Operations and Properties

Matrix Transpose

- Denoted as $\mathbf{A}^{\mathrm{T}}$
- Defined as

$$\mathbf{(A^{T})_{i,j} = A_{j,I}}$$

- Mirror image across the (main) diagonal of a matrix
  - Main diagonal -> running down from upper left to the bottom right

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} & A_{1,3} \\ A_{2,1} & A_{2,2} & A_{2,3} \\ A_{3,1} & A_{3,2} & A_{3,3} \end{bmatrix} \Rightarrow A^{T} = \begin{bmatrix} A_{1,1} & A_{2,1} & A_{3,1} \\ A_{1,2} & A_{2,2} & A_{3,2} \\ A_{1,3} & A_{2,3} & A_{3,3} \end{bmatrix}$$

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \\ A_{3,1} & A_{3,2} \end{bmatrix} \Rightarrow A^{T} = \begin{bmatrix} A_{1,1} & A_{2,1} & A_{3,1} \\ A_{1,2} & A_{2,2} & A_{3,2} \end{bmatrix}$$

Matrix Transpose Examples

- Transpose a Vector

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x_1} \\ \boldsymbol{x_2} \\ \vdots \\ \boldsymbol{x_n} \end{bmatrix} \Rightarrow \boldsymbol{x^T} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^T$$

- Transpose a scalar
  - Scalar → Matrix with one element

$$a = a^T$$

Matrix Inverse

- Not all matrix has an inverse
- If A$\in$ $\mathbb{R}^{n\times n}$ matrix, and if it has an inverse, then

$A(A^{-1}) = (A^{-1})A = I$

- Matrices that don't have an inverse are "singular" or "degenerate"

Matrix Addition

- To perform addition, two matrices must have same dimensions
  - Same dimension = same number of rows and columns

- **C= A + B**

- $\begin{bmatrix} 12 & 34 \\ -123 & 15 \\ 14 & 10 \end{bmatrix} + \begin{bmatrix} 2 & 3 \\ -1 & 15 \\ 10 & 10 \end{bmatrix} = \begin{bmatrix} 24 & 37 \\ ? & ? \\ ? & ? \end{bmatrix}$

$\mathbf{A} = \mathbb{R}^{3\times2}$ $\mathbf{B} = \mathbb{R}^{3\times2}$ $\mathbf{C} = \mathbb{R}^{3\times2}$

Matrix Addition

- To perform addition, two matrices must have same dimensions
  - Same dimension = same number of rows and columns

  - **C= A + B**

  - $$\begin{bmatrix} 12 & 34 \\ -123 & 15 \\ 14 & 10 \end{bmatrix} + \begin{bmatrix} 2 & 3 \\ -1 & 15 \\ 10 & 10 \end{bmatrix} = \begin{bmatrix} 24 & 37 \\ -124 & 30 \\ 24 & 20 \end{bmatrix}$$

$\mathbf{A} = \mathbb{R}^{3\times 2}$ $\mathbf{B} = \mathbb{R}^{3\times 2}$ $\mathbf{C} = \mathbb{R}^{3\times 2}$

Matrix Addition

- **C= A + B =?**
- $\begin{bmatrix} 12 & 34 \\ -123 & 15 \\ 14 & 10 \end{bmatrix} + \begin{bmatrix} 2 & 3 \\ -1 & 15 \end{bmatrix} = \ ?$

$\mathbf{A} = \ \mathbb{R}^{3\times 2}$ $\mathbf{B} = \ \mathbb{R}^{2\times 2}$

Matrix Addition

- **C= A + B =?**
- $\begin{bmatrix} 12 & 34 \\ -123 & 15 \\ 14 & 10 \end{bmatrix} + \begin{bmatrix} 2 & 3 \\ -1 & 15 \end{bmatrix} = \boldsymbol{ERROR}, \text{dimension mismatch}$

$\mathbf{A} = \mathbb{R}^{3\times2}$ $\mathbf{B} = \mathbb{R}^{2\times2}$

Matrix and a Scalar Addition

- $\mathbf{A} \in \mathbb{R}^{m \times n}$
- $\mathbf{C} = \mathbf{A} + k$
  - $\mathbf{C}_{i,j} = \mathbf{A}_{i,j} + k$

```python
import numpy as np


A = np.array([[1, 2, 3],
              [6, 7, 9]])

k = -5

C = A + k

print(C)
```

Matrix and a Scalar Addition

- $\mathbf{A} \in \mathbb{R}^{m \times n}$
- $\mathbf{C} = \mathbf{A} + k$
    - $\mathbf{C}_{i,j} = \mathbf{A}_{i,j} + k$

```python
import numpy as np


A = np.array([[1, 2, 3],
              [6, 7, 9]])

k = -5

C = A + k

print(C)
```

```
[[-4 -3 -2]
 [ 1  2  4]]
```

Matrix Multiplication

- For **C = AB**,
  - Number of columns in A = number of rows of B (**MUST**)
  - If **A** is of shape *mxn* and **B** is of shape *nxp* then matrix product **C** is of shape *mxp*

$$C = AB \Rightarrow C_{i,j} = \sum_k A_{i,k} B_{k,j}$$

Matrix Multiplication: Matrix and a Vector

- $\begin{bmatrix} 1 & 4 & 2 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$

$\mathbf{A} = \mathbb{R}^{2\times 3}$ $\mathbf{B} = \mathbb{R}^{3\times 1}$ $\mathbf{C} = \mathbb{R}^{2\times 1}$

- 1x 1 + 4x 0+ 2x 3 = 7 $\begin{bmatrix} \boxed{1 \quad 4 \quad 2} \\ 3 \quad 0 \quad 1 \end{bmatrix} \boxed{\begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}}$

- 3x1 + 0x0 + 1x3 = 6 $\begin{bmatrix} 1 \quad 4 \quad 2 \\ \boxed{3 \quad 0 \quad 1} \end{bmatrix} \boxed{\begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}}$

Dot Product

$\mathbf{a}\cdot\mathbf{b} = a_1b_1 + a_2b_2 + a_3b_3$

Matrix Multiplication: Matrix and a Matrix

- $\begin{bmatrix} 1 & 4 & 2 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 3 & 0 \end{bmatrix} = \begin{bmatrix} 7 & 6 \\ 6 & 6 \end{bmatrix}$

  $\mathbf{A} = \mathbb{R}^{2\times 3}$ $\mathbf{B} = \mathbb{R}^{3\times 2}$ $\mathbf{C} = \mathbb{R}^{2\times 2}$

- 1x 1 + 4x 0+ 2x 3 = 7
- 3x1 + 0x0 + 1x3 = 6

- 1x2 + 4x1 + 2x0 = 6
- 3x2 + 0x1 + 1x0 = 6

Matrix Product vs Elementwise Multiplication

- Matrix product **is not** elementwise multiplication
  - Hadamard product, denoted as $A \odot B$

$$\begin{bmatrix} 2 & 3 & 1 \\ 0 & 8 & -2 \end{bmatrix} \circ \begin{bmatrix} 3 & 1 & 4 \\ 7 & 9 & 5 \end{bmatrix} = \begin{bmatrix} 2 \times 3 & 3 \times 1 & 1 \times 4 \\ 0 \times 7 & 8 \times 9 & -2 \times 5 \end{bmatrix} = \begin{bmatrix} 6 & 3 & 4 \\ 0 & 72 & -10 \end{bmatrix}$$

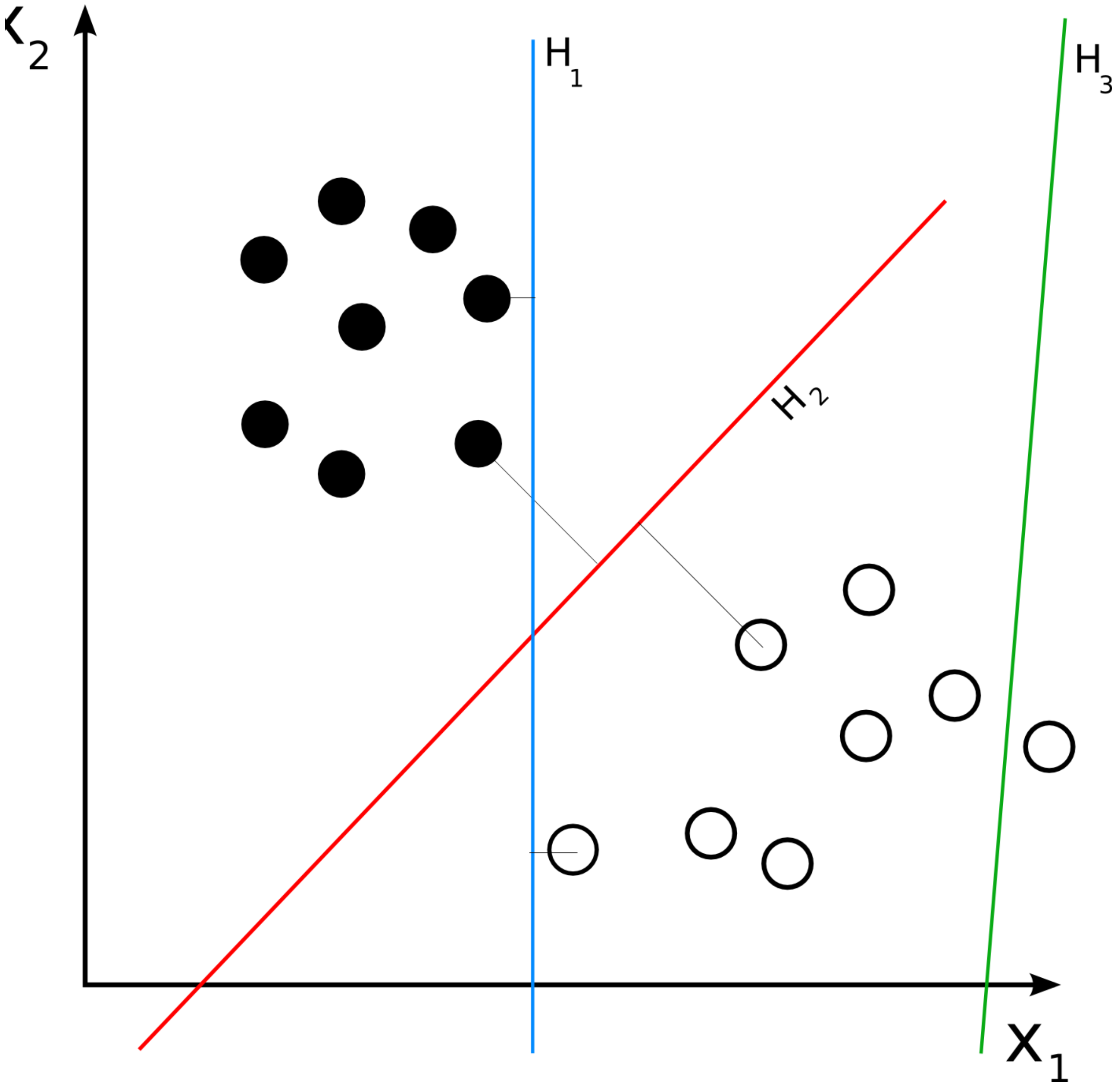
Example Source: Wikipedia

Matrix Product Properties

- *Distributivity over addition:* **A(B+C)=AB+AC**
- *Associativity:* **A(BC)=(AB)C**
- *NOT commutative:* **AB=BA** is not always true
- *Dot product between vectors is commutative:* $x^T y = y^T x$
- *Transpose of a matrix product has a simple form:* $\mathbf{(AB)^T = B^T A^T}$

Linear Classifier

- The simplest ML model
- Makes a *classification* decision based on the value of ***a linear combination*** of the characteristics (features).
- Black and white circles are different labels. H_1, H_2, … represent different *decision boundaries* i.e. linear functions that best map the classification process.
    - ***Goal:*** find the best linear function that has highest accuracy



Linear Classifier (cntd.)

- Mathematically represented as

$$y = \boldsymbol{W}\boldsymbol{x}^{\boldsymbol{T}} + b$$

where y → labels (vector)

W → model parameter matrix

x → feature vector

b → bias term (scalar)

- Very similar in the mathematical representation of a line

$$y = mx + c$$

→Hence the term ***linear classifier***

Linear Classifier Example

define a **score function**

class scores

Slide Credit:Dr. Fei Fei Li and Andrej Karpathy

Linear Classifier Example

- Score function $f = (W, x_i, b)$

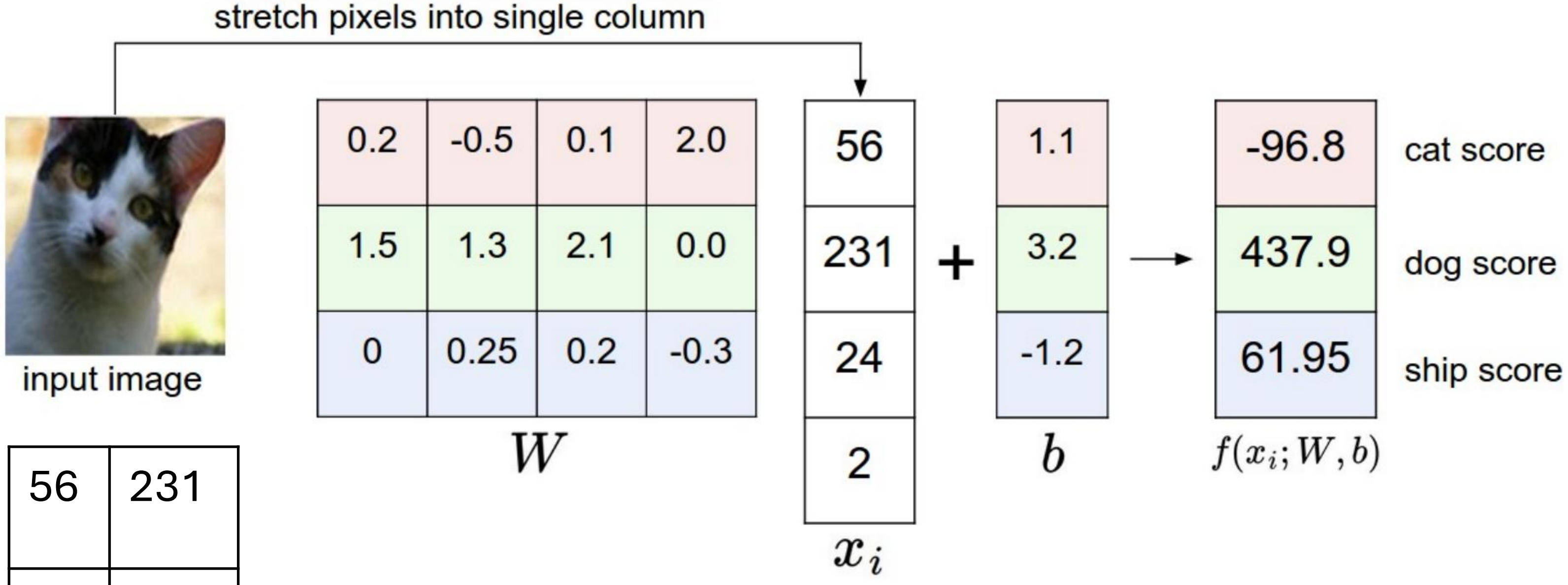
- $x_{i\,=}$ input image from CIFAR-10 example,
  - CIFAR-10 images = 32 x 32 x 3 = 3072 and 10 classes

Assume
- Dimension of $x_i = 4$
- 3 classes

$f = (W, x_i, b) = W x_i + b$

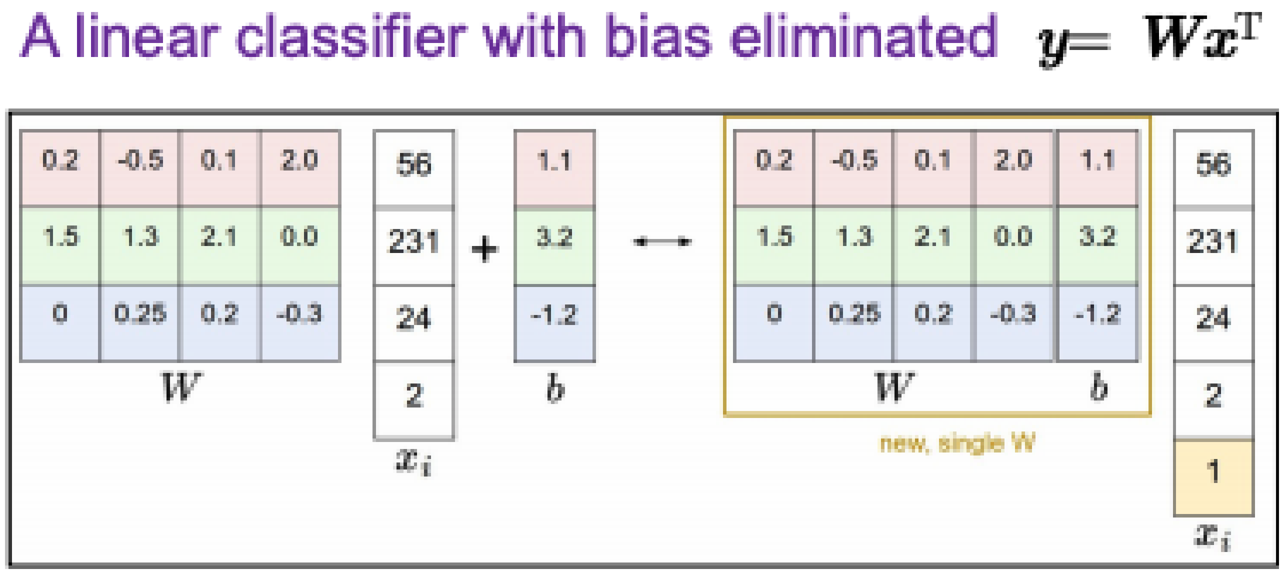
$$\mathbf{W} = 3 * 4$$
$$\mathbf{x_i} = 4 * 1$$
$$\mathbf{b} = 3 * 1$$
$$f = 3 * 1$$

Linear Classifier Example



56	231
24	2

Linear Classifier Example



Linear Transformation

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$$

- where $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ and $\boldsymbol{b} \in \mathbb{R}^{n}$
- More explicitly

$$\begin{aligned} A_{11}x_1 + A_{12}x_2 + \dots + A_{1n}x_n &= b_1 \\ A_{21}x_1 + A_{22}x_2 + \dots + A_{2n}x_n &= b_2 \\ \\ A_{n1}x_1 + A_{m2}x_2 + \dots + A_{n,n}x_n &= b_n \end{aligned}$$

n equations in n unknowns

$$\boldsymbol{A} = \begin{bmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & \vdots & \vdots \\ A_{n,1} & \cdots & A_{nn} \end{bmatrix} \qquad \boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \qquad \boldsymbol{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

n x n $\qquad$ n x 1 $\qquad$ n x 1

Can view A as a *linear transformation* of vector $\boldsymbol{x}$ to vector $\boldsymbol{b}$

Linear Transformation

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$$

– where $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ and $\boldsymbol{b} \in \mathbb{R}^{n}$

– More explicitly

$$\boxed{\begin{aligned} A_{11}x_1 + A_{12}x_2 + \ldots + A_{1n}x_n &= b_1 \\ A_{21}x_1 + A_{22}x_2 + \ldots + A_{2n}x_n &= b_2 \\ \\ A_{n1}x_1 + A_{m2}x_2 + \ldots + A_{n,n}x_n &= b_n \end{aligned}}$$

n equations in
n unknowns

$$\boxed{\boldsymbol{A} = \begin{bmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & \vdots & \vdots \\ A_{n,1} & \cdots & A_{nn} \end{bmatrix} \quad \boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \boldsymbol{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}}$$

n x n $\quad$ n x 1 $\quad$ n x 1

Can view A as a *linear transformation* of vector $\boldsymbol{x}$ to vector $\boldsymbol{b}$

How to solve this?

Linear Transformation(cntd.)

- Solving $\mathbf{A}\boldsymbol{x}=\boldsymbol{b}$:

$$\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$$

$$\mathbf{A}^{-1}\mathbf{A}\boldsymbol{x} = \mathbf{A}^{-1}\boldsymbol{b}$$

$$\mathbf{I}_n\boldsymbol{x} = \mathbf{A}^{-1}\boldsymbol{b}$$

$$\boldsymbol{x} = \mathbf{A}^{-1}\boldsymbol{b}$$

Linear Transformation (cntd.)

- If $\mathbf{A}^{-1}$ exists there are several methods for finding it

- **Alternative:** Use *Gaussian elimination* and back-substitution.

Gaussian Elimination

- Goal: Transform the matrix **A** into an upper triangular matrix

- Steps:

1. Given a linear system in matrix form, $A\mathbf{x} = \mathbf{b}$, first write down the corresponding augmented matrix: $[A|\mathbf{b}]$

2. Perform sequence of row operations

   Type 1. Interchange any two rows.

   Type 2. Multiply a row by a nonzero constant.

   Type 3. Add a multiple of one row to another row.

Gaussian Elimination Example

$$
\begin{aligned}
2x \quad - 2y \qquad &= -6 \\
x \quad - 2y + z \ &= 1 \\
3y - 2z \ &= -5
\end{aligned}
$$

- Augmented matrix
- $\begin{bmatrix} 2 & -2 & 0 & -6 \\ 1 & -1 & 1 & 1 \\ 0 & 3 & -2 & -5 \end{bmatrix}$

Gaussian Elimination Example (cntd.)

- $\begin{bmatrix} 2 & -2 & 0 & -6 \\ 1 & -1 & 1 & 1 \\ 0 & 3 & -2 & -5 \end{bmatrix} \xrightarrow{\text{r1 * ½}} \begin{bmatrix} 1 & -1 & 0 & -3 \\ 1 & -1 & 1 & 1 \\ 0 & 3 & -2 & -5 \end{bmatrix}$

- $\begin{bmatrix} 1 & -1 & 0 & -3 \\ 1 & -1 & 1 & 1 \\ 0 & 3 & -2 & -5 \end{bmatrix} \xrightarrow{\text{-r1+ r2}} \begin{bmatrix} 1 & -1 & 0 & -3 \\ 0 & 0 & 1 & 4 \\ 0 & 3 & -2 & -5 \end{bmatrix}$

Gaussian Elimination Example (cntd.)

- $\begin{bmatrix} 1 & -1 & 0 & -3 \\ 0 & 0 & 1 & 4 \\ 0 & 3 & -2 & -5 \end{bmatrix} \xrightarrow{r2 \leftrightarrow r3} \begin{bmatrix} 1 & -1 & 0 & -3 \\ 0 & 3 & -2 & -5 \\ 0 & 0 & 1 & 4 \end{bmatrix}$

- From third row, z = 4.
- Back-substituting z= 4 into the second row gives y = 1
- Back-substitution of y= 1into the first row yields x = −2

Norms

- Measures the size of a vector

- The norm has to satisfy three properties

$$f\left(\boldsymbol{x}\right)=0 \Rightarrow \boldsymbol{x}=0$$

$$f(\boldsymbol{x}+\boldsymbol{y}) \leq f\left(\boldsymbol{x}\right)+f\left(\boldsymbol{y}\right) \quad \text{Triangle Inequality}$$

$$\forall \alpha \in R \quad f\left(\alpha \boldsymbol{x}\right)=\left|\alpha\right| f\left(\boldsymbol{x}\right)$$

Norms: L1 and L2 Norms

L1: sum of the absolute value of the entries in the vector

$$\|x\|_1 = \sum |x_i|$$

L2: square root of the sum of the squares of entries of the vector.

$$\|x\|_2 = \sqrt{\sum x_i^2}$$

Lp : pth root of the sum of the entries of the vector raised to the pth power

$$\|x\|_p = (\sum |x_i|^p)^{1/p}$$

Frequently used: square of the l2 norm

$$\|x\|_2^2 = \sum x_i^2$$

Example Norms

- A =(cos(x),sin(x)) Norm= ?

- $$\mathbf{a} = \begin{bmatrix} 2 \\ 1 \\ 7 \\ -2 \end{bmatrix}$$ Norm= ?

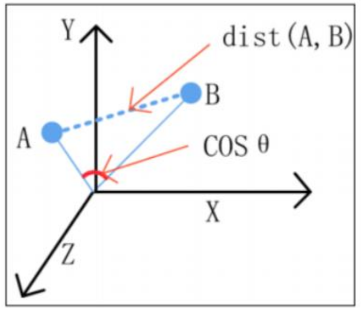
Example Norms (cntd.)

- A =(cos(x),sin(x))
- Norm = sqrt($cos^2x + sin^2$x) = 1
- $$\mathbf{a} = \begin{bmatrix} 2 \\ 1 \\ 7 \\ -2 \end{bmatrix}$$

Norm = sqrt($2^2 + 1^2 + 7^2$+ $(-2)^2$) = sqrt(58)

“Special” Vectors

- Unit vector:
  - A vector with unit norm: $L^2(x) = 1$
- Orthogonal vectors:
  - Vectors x and y are orthogonal if $x^T y = 0$
    - i.e. if the vectors have non-zero norm, they are at 90 degrees to each other
- Orthonormal vector:
  - Vectors are orthogonal and have unit norm
- Dot product of two vectors: $\boxed{\boldsymbol{x}^T\boldsymbol{y} \Rightarrow ||\boldsymbol{x}||_2 ||\boldsymbol{y}||_2 \cos\boldsymbol{\theta}}$



Distance between two vectors $(\boldsymbol{v},\boldsymbol{w})$

– $\mathrm{dist}(\boldsymbol{v},\boldsymbol{w}) = ||\boldsymbol{v}-\boldsymbol{w}||$

$$= \sqrt{(v_1 - w_1)^2 + .. + (v_n - w_n)^2}$$

Optional Reading

- Eigen Values and Eigenvectors
- Singular value decomposition
- Handout provided