Machine Learning 2007: Lecture 2

Instructor: Tim van Erven (Tim.van.Erven@cwi.nl)
Website: `www.cwi.nl/~erven/teaching/0708/ml/`

September 13, 2007

Overview

- **Organisational Matters**
- This Lecture versus Mitchell
- Vectors and Matrices
  - Scalars, Vectors and Matrices
  - Addition
  - Multiplication by a Scalar
  - The Transpose
  - Multiplying Vectors or Matrices
  - The Identity Matrix
  - The Matrix Inverse
- Data Representation Using Vectors

Organisational Matters

- Please register on Blackboard:
  Machine Learning (2007-2008)_1
- Final exam: December 20, 18.30 – 21.15
- Homework Exercises 1 moved to this week. I will make an alternative version available for students who have not seen vectors and matrices before.

Overview

- Organisational Matters
- **This Lecture versus Mitchell**
- Vectors and Matrices
  - Scalars, Vectors and Matrices
  - Addition
  - Multiplication by a Scalar
  - The Transpose
  - Multiplying Vectors or Matrices
  - The Identity Matrix
  - The Matrix Inverse
- Data Representation Using Vectors

This Lecture versus Mitchell

Mitchell

- Still Chapter 1 and Chapter 2 up to section 2.2. (Be patient, we will go faster soon enough.)

This Lecture

- Vectors and matrices are not in Mitchell.
- There is no explicit discussion on data representation in Mitchell.

Overview

- Organisational Matters
- This Lecture versus Mitchell
- Vectors and Matrices
  - **Scalars, Vectors and Matrices**
  - Addition
  - Multiplication by a Scalar
  - The Transpose
  - Multiplying Vectors or Matrices
  - The Identity Matrix
  - The Matrix Inverse
- Data Representation Using Vectors

Scalars and Vectors

Scalars:

A scalar α is just an ordinary number (element of $\mathbb{R}$).

- For example: $x = 10$.

Vectors:

An n-dimensional vector $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ is a list of n numbers.

- For example, $\mathbf{x} = \begin{pmatrix} 3 \\ -4/7 \\ \pi \\ 10 \end{pmatrix}$.
- Note the convention of writing the list vertically.

The Vector Space $\mathbb{R}^n$

The set of all n-dimensional vectors is defined as:

$$\mathbb{R}^n = \left\{ \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \middle| x_1 \in \mathbb{R}, \ldots, x_n \in \mathbb{R} \right\}.$$

- Such spaces are called **vector spaces**.
- Geometrically, $\mathbb{R}^1 = \mathbb{R}$ is a line.
- $\mathbb{R}^2$ is a plane.
- $\mathbb{R}^3$ is the 3-dimensional space.
- $\mathbb{R}^n$ is the n-dimensional space.

Matrices

Definition:

An $m \times n$ matrix A with elements a_{ij} is an array of numbers:

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

- In a_{ij}: i indicates the row and j indicates the column.
- An $m \times 1$ matrix is an m-dimensional vector.

Example:

$$A = \begin{pmatrix} 10 & -3 & 1 & 7 \\ \pi & 6 & -1/9 & 2 \\ 0 & 0 & 1 & 2 \end{pmatrix}$$

Overview

- Organisational Matters
- This Lecture versus Mitchell
- Vectors and Matrices
  - Scalars, Vectors and Matrices
  - **Addition**
  - Multiplication by a Scalar
  - The Transpose
  - Multiplying Vectors or Matrices
  - The Identity Matrix
  - The Matrix Inverse
- Data Representation Using Vectors

Adding Vectors

Definition:

For any two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$

$$\mathbf{x} + \mathbf{y} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}.$$

- You can not add vectors of different dimensionality.

Example:

$$\mathbf{x} + \mathbf{y} = \begin{pmatrix} 3 \\ 10 \\ -4/7 \\ \pi \\ 0 \end{pmatrix} + \begin{pmatrix} 6 \\ -5 \\ 4 \\ -3 \\ 2 \end{pmatrix} = \begin{pmatrix} 9 \\ 5 \\ 24/7 \\ \pi - 3 \\ 2 \end{pmatrix}$$

Adding Matrices

Definition:

For any two $m \times n$ matrices A and B,

$$A + B = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} + \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mn} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{pmatrix}.$$

- You can not add matrices of different dimensionality.

Adding Matrices

Example:

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix} + \begin{pmatrix} -1 & 1 & -1 \\ 1 & \pi & 1 \\ -1 & 1 & -1 \\ 0 & 6 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 3 & 2 \\ 5 & 5+\pi & 7 \\ 6 & 9 & 8 \\ 10 & 17 & 12 \end{pmatrix}$$

But this is not defined:

$$\begin{pmatrix} 1 & 1 & 2 \\ 3 & 5 & 8 \end{pmatrix} + \begin{pmatrix} 13 & 21 \\ 34 & 55 \\ 89 & 144 \end{pmatrix} = \text{?}$$

Overview

- Organisational Matters
- This Lecture versus Mitchell
- Vectors and Matrices
    - Scalars, Vectors and Matrices
    - Addition
    - **Multiplication by a Scalar**
    - The Transpose
    - Multiplying Vectors or Matrices
    - The Identity Matrix
    - The Matrix Inverse
- Data Representation Using Vectors

Multiplying a Vector by a Scalar

Definition:

For any vector $\mathbf{x} \in \mathbb{R}^n$ and scalar $\alpha \in \mathbb{R}$

$$\alpha\mathbf{x} = \alpha \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \alpha x_1 \\ \vdots \\ \alpha x_n \end{pmatrix}.$$

Example:

$$2\mathbf{x} = 2 \begin{pmatrix} 3 \\ 10 \\ -4/7 \\ \pi \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 6 \\ 20 \\ -8/7 \\ 2\pi \\ 0 \\ -2 \end{pmatrix}$$

Multiplying a Matrix by a Scalar

Definition:

For any $m \times n$ matrix A and scalar $\alpha \in \mathbb{R}$

$$\alpha A = \alpha \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} = \begin{pmatrix} \alpha a_{11} & \cdots & \alpha a_{1n} \\ \vdots & \ddots & \vdots \\ \alpha a_{m1} & \cdots & \alpha a_{mn} \end{pmatrix}.$$

Example:

$$-2 \begin{pmatrix} 3 & -1 & -9 \\ -4 & 5 & 4 \end{pmatrix} = \begin{pmatrix} -6 & 2 & 18 \\ 8 & -10 & -8 \end{pmatrix}$$

Overview

- Organisational Matters
- This Lecture versus Mitchell
- Vectors and Matrices
  - Scalars, Vectors and Matrices
  - Addition
  - Multiplication by a Scalar
  - **The Transpose**
  - Multiplying Vectors or Matrices
  - The Identity Matrix
  - The Matrix Inverse
- Data Representation Using Vectors

The Matrix and Vector Transpose

Matrix Transpose:

If A is an $m \times n$ matrix with elements a_{ij}, then its transpose $A^\top$ is an $n \times m$ matrix with elements b_{ij} such that $b_{ij} = a_{ji}$.

Example:

$$\begin{pmatrix} 1 & -8 & 2 \\ 3 & 5 & 1 \end{pmatrix}^\top = \begin{pmatrix} 1 & 3 \\ -8 & 5 \\ 2 & 1 \end{pmatrix}$$

Vector Transpose:

An m-dimensional vector is a $m \times 1$ matrix. Therefore the vector transpose is a special case of the matrix transpose. For example,

$$\begin{pmatrix} 9 \\ -3 \\ -1 \end{pmatrix}^\top = \begin{pmatrix} 9 & -3 & -1 \end{pmatrix} \qquad \begin{pmatrix} 9 & -3 & -1 \end{pmatrix}^\top = \begin{pmatrix} 9 \\ -3 \\ -1 \end{pmatrix}$$

Overview

- Organisational Matters
- This Lecture versus Mitchell
- Vectors and Matrices
  - Scalars, Vectors and Matrices
  - Addition
  - Multiplication by a Scalar
  - The Transpose
  - **Multiplying Vectors or Matrices**
  - The Identity Matrix
  - The Matrix Inverse
- Data Representation Using Vectors

Multiplying Vectors

Definition:

If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ are n-dimensional vectors, then their **inner product**, denoted $\langle \mathbf{x}, \mathbf{y} \rangle$, is defined as

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} x_i y_i = x_1 y_1 + \ldots + x_n y_n$$

Example:

$$\langle \begin{pmatrix} 9 \\ 5 \\ 1 \end{pmatrix}, \begin{pmatrix} -3 \\ 2 \\ 11 \end{pmatrix} \rangle = 9 \cdot -3 + 5 \cdot 2 + 1 \cdot 11 = -6$$

Multiplying Matrices

Some Notation:

We may view a matrix as a collection of vectors:

$$A = \begin{pmatrix} \text{——} & \mathbf{a}_1^\top & \text{——} \\ & \vdots & \\ \text{——} & \mathbf{a}_m^\top & \text{——} \end{pmatrix} \qquad B = \begin{pmatrix} | & & | \\ \mathbf{b}_1 & \cdots & \mathbf{b}_n \\ | & & | \end{pmatrix}$$

Matrix Product:

If A is an $m \times k$ matrix and B is a $k \times n$ matrix, then their product AB is the $m \times n$ matrix with elements c_{ij} such that

$$c_{ij} = \langle \mathbf{a}_i, \mathbf{b}_j \rangle$$

- Note that $\langle \mathbf{a}_i, \mathbf{b}_j \rangle = \sum_{l=1}^{k} a_{il} b_{lj}$.

Multiplying Matrices

Example:

$$\begin{pmatrix} 3 & -1 \\ -4 & 5 \\ 4 & -9 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 0 \end{pmatrix} = \begin{pmatrix} \langle \left(\begin{smallmatrix} 3 \\ -1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 \\ 3 \end{smallmatrix}\right) \rangle & \langle \left(\begin{smallmatrix} 3 \\ -1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 2 \\ 0 \end{smallmatrix}\right) \rangle \\ \langle \left(\begin{smallmatrix} -4 \\ 5 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 \\ 3 \end{smallmatrix}\right) \rangle & \langle \left(\begin{smallmatrix} -4 \\ 5 \end{smallmatrix}\right), \left(\begin{smallmatrix} 2 \\ 0 \end{smallmatrix}\right) \rangle \\ \langle \left(\begin{smallmatrix} 4 \\ -9 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 \\ 3 \end{smallmatrix}\right) \rangle & \langle \left(\begin{smallmatrix} 4 \\ -9 \end{smallmatrix}\right), \left(\begin{smallmatrix} 2 \\ 0 \end{smallmatrix}\right) \rangle \end{pmatrix}$$

$$= \begin{pmatrix} 3 - 3 & 6 - 0 \\ -4 + 15 & -8 + 0 \\ 4 - 27 & 8 - 0 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 6 \\ 11 & -8 \\ -23 & 8 \end{pmatrix}$$

Multiplying a Matrix and a Vector

A Special Case of Matrix Multiplication:

- Recall that a k-dimensional vector is a $k \times 1$ matrix.
- Hence if A is an $m \times k$ matrix and $\mathbf{x}$ is a k-dimensional vector, then the product $A\mathbf{x}$ is an $m \times 1$ matrix, which is a m-dimensional vector.

Example:

$$\begin{pmatrix} 3 & -1 \\ -4 & 5 \\ 4 & -9 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} \langle \left(\begin{smallmatrix} 3 \\ -1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 \\ 3 \end{smallmatrix}\right) \rangle \\ \langle \left(\begin{smallmatrix} -4 \\ 5 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 \\ 3 \end{smallmatrix}\right) \rangle \\ \langle \left(\begin{smallmatrix} 4 \\ -9 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 \\ 3 \end{smallmatrix}\right) \rangle \end{pmatrix}$$

$$= \begin{pmatrix} 3 - 3 \\ -4 + 15 \\ 4 - 27 \end{pmatrix} = \begin{pmatrix} 0 \\ 11 \\ -23 \end{pmatrix}$$

Overview

- Organisational Matters
- This Lecture versus Mitchell
- Vectors and Matrices
  - Scalars, Vectors and Matrices
  - Addition
  - Multiplication by a Scalar
  - The Transpose
  - Multiplying Vectors or Matrices
  - **The Identity Matrix**
  - The Matrix Inverse
- Data Representation Using Vectors

The Identity Matrix

The Identity Matrix:

- The $n \times n$ identity matrix I satisfies $I\mathbf{x} = \mathbf{x}$ for all vectors $\mathbf{x}$.
- It has 1s on the diagonal and 0s everywhere else.

Example:

$$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Overview

- Organisational Matters
- This Lecture versus Mitchell
- Vectors and Matrices
  - Scalars, Vectors and Matrices
  - Addition
  - Multiplication by a Scalar
  - The Transpose
  - Multiplying Vectors or Matrices
  - The Identity Matrix
  - **The Matrix Inverse**
- Data Representation Using Vectors

Matrix Inverse

Definition:

Suppose A is an $n \times n$ matrix. Then the matrix inverse A^{-1} (if it exists) is the $n \times n$ matrix such that

$$A^{-1}A = I \quad \text{and} \quad AA^{-1} = I.$$

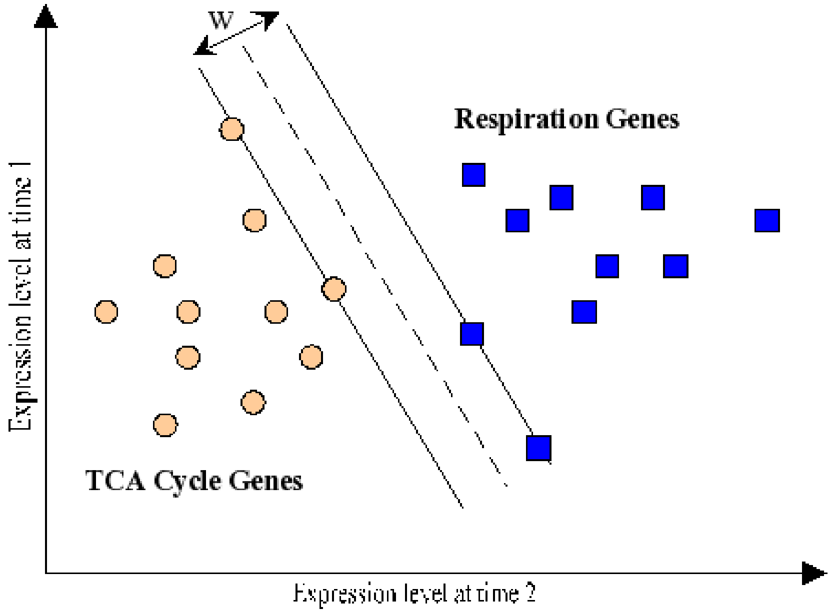
Example:

$$A = \begin{pmatrix} 2 & 3 & 1 \\ 2 & 6 & 1 \\ -2 & 3 & 0 \end{pmatrix} \quad A^{-1} = \begin{pmatrix} 1/2 & -1/2 & -1/2 \\ -1/3 & 1/3 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

Overview

- Organisational Matters
- This Lecture versus Mitchell
- Vectors and Matrices
  - Scalars, Vectors and Matrices
  - Addition
  - Multiplication by a Scalar
  - The Transpose
  - Multiplying Vectors or Matrices
  - The Identity Matrix
  - The Matrix Inverse
- **Data Representation Using Vectors**

Classifying Genes by Gene Expression

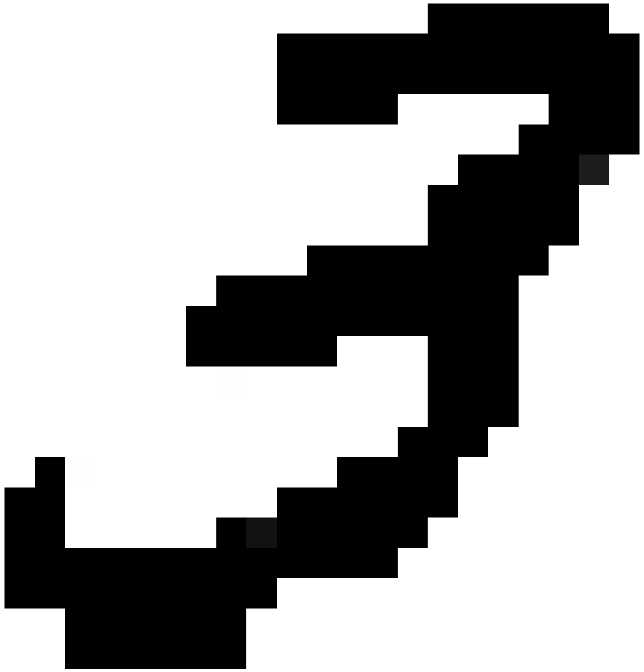


$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

where $x_1, x_2 \in \mathbb{R}$ are the expression levels at times 1 and 2, respectively.

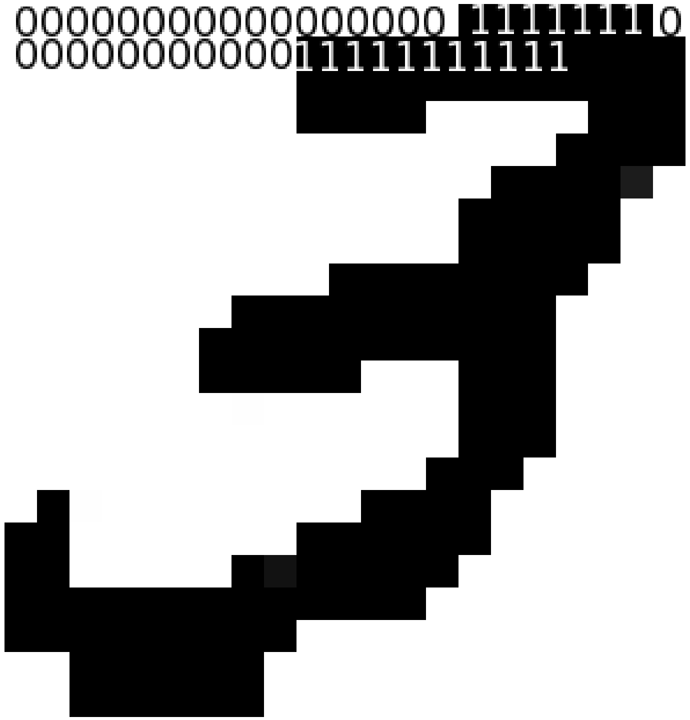
Handwritten Digits

Consider the problem of classifying handwritten digits again [LeCun et al., 1998]. How can we represent such digits as vectors?



Handwritten Digits

Consider the problem of classifying handwritten digits again [LeCun et al., 1998]. How can we represent such digits as vectors? Concatenate rows.



$$\Longrightarrow \mathbf{x} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 1 \\ 0 \\ 0 \\ 0 \\ \vdots \end{pmatrix}$$

Checkers Board Features

Don’t represent the entire board, but only aspects of it (Mitchell):

$$\text{Checkers board} \Longrightarrow \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix}$$

Feature	Meaning
x_1	the number of black pieces on the board
x_2	the number of red pieces on the board
x_3	the number of black kings on the board
x_4	the number of red kings on the board
x_5	the number of black pieces threatened by red
x_6	the number of red pieces threatened by black

EnjoySport 1

One Way to Do It:

Attribute	Sky			AirTemp	
Value	Sunny	Cloudy	Rainy	Warm	Cold
Encoding	1	2	3	1	2

$$\text{Sunny, Warm} \Longrightarrow \mathbf{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\text{Rainy, Cold} \Longrightarrow \mathbf{x} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

$$\text{Sunny, Cold} \Longrightarrow \mathbf{x} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

- The difference of for example $\begin{pmatrix} 3 \\ 2 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ has no meaning.

EnjoySport 2

Another Way to Do It:

Attribute	Sky			AirTemp	
Value	Sunny	Cloudy	Rainy	Warm	Cold
Encoding	$\begin{pmatrix}1\\0\\0\end{pmatrix}$	$\begin{pmatrix}0\\1\\0\end{pmatrix}$	$\begin{pmatrix}0\\0\\1\end{pmatrix}$	$\begin{pmatrix}1\\0\end{pmatrix}$	$\begin{pmatrix}0\\1\end{pmatrix}$

$$\text{Sunny, Warm} \Longrightarrow \mathbf{x} = \begin{pmatrix}1\\0\\0\\1\\0\end{pmatrix} \qquad \text{Rainy, Cold} \Longrightarrow \mathbf{x} = \begin{pmatrix}0\\0\\1\\0\\1\end{pmatrix}$$

- The number of non-zero entries in the difference between two vectors is twice the number of attributes that differ.

References

- Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, ”Gradient-Based Learning Applied to Document Recognition,” Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, Nov. 1998.
- C. Geijsel, and W. Hoffmann, “Lineaire Algebra voor Informatiewetenschappen,” Universiteit van Amsterdam, Dec. 2000