6.1 Data Matrices

Let $x^{(1)}, x^{(2)}, \ldots, x^{(l)}$ be $l$ input vectors, each $x^{(i)} \in \mathbb{R}^n$. Convention: arrange $x^{(i)}$ as columns of a matrix (called the "data matrix")

$$X = \begin{pmatrix} x^{(1)} & x^{(2)} & \ldots & x^{(l)} \end{pmatrix} \in \mathbb{R}^{n \times l}$$

Example 6.1. MicroArrays – gene expression profiling. Columns are different experiments, e.g. rows are different genes, Entries are expression levels ($a_{ij}$ states expression of gene $i$ in experiment $j$).

Example 6.2. EEG. Each column is one electrode, rows correspond to different time steps. Entries are electric field strength.

Example 6.3. Neural Recordings. Each column is one neuron, rows correspond to different time steps. Entries are potentials or (binarize) flag if neuron is spiking.

Evaluating the same function multiple times with different inputs

Data matrices are particularly useful if we want to apply a fixed linear function (e.g. averaging, or computing an expectation) to multiple data vectors:

- fixed $f : \mathbb{R}^n \to \mathbb{R}^m$, $f(x) = Ax$ with $A \in \mathbb{R}^{m \times n}$.
- we want to compute $f(x^{(i)})$ for multiple vectors, $x^{(1)}, \ldots, x^{(k)}$. 

---

33
We can arrange the output vectors in matrix form as well
\[
Y = \begin{pmatrix} f(x^{(1)}) & f(x^{(2)}) & \ldots & f(x^{(l)}) \end{pmatrix} \in \mathbb{R}^{m \times l}
\] (6.2)

What’s the relation between \(X\) and \(Y\)?

\[
y_{ij} = [f(x^{(j)})]_i = [Ax^{(j)}]_i = \sum_{k=1}^{l} a_{ik} x_{kj} = \sum_{k=1}^{l} a_{ik} x_{kj}
\] (6.3)

Note: this is the same rule as in matrix multiplication

\[
Y = AX
\] (6.4)

**Observation.** Matrices serve simultaneously (at least) two purposes:
- they can store the 'instructions' for a linear function,
- they can store data.

The rules how to handle both cases are the same, we do not have to distinguish between both situations.

More Examples of Matrices that Represent Data

**Example 6.4. Genetic Distances** For \(n\) organisms we have their genomes \(g_1, \ldots, g_n\). We can compute the similarities between any pair of organisms, \(sim(g_i, g_j)\). These can naturally be arrange as a \(n \times n\) matrix: \(A \in \mathbb{R}^{n \times n}\) with \(a_{ij} = sim(g_i, g_j)\).

**Example 6.5. (Facebook)** Social networks make a lot of use of matrices. Assume Facebook has \(n\) users (real life: \(n \geq 1,000,000,000\) since Sep. 2012). We can encode **who-is-a-friend-of-who** in a friendship matrix: \(A \in \mathbb{R}^{n \times n}\) with

\[
a_{ij} = \begin{cases} 
1 & \text{if person } i \text{ and person } j \text{ are "friends"} \\
0 & \text{otherwise.}
\end{cases}
\]

We obtain information about the friendship network from \(A\):
• $b_i = \sum_j a_{ij}$ is number of friends that person $i$ has. We can compute this for all people jointly as

$$b = A1_n, \text{ where } 1_n = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^n.$$

• $c := \frac{1}{n} \sum_{i=1}^n b_i$ is the average number of friends that people in Facebook have. (real life: $c \approx 140$)

The special form a $A$ allows more: Let’s look at $B = AA = A^2$:

$$b_{ij} = \sum_{k=1}^n a_{ik}a_{kj} \quad (6.5)$$

• $a_{ik}a_{kj} = 1$ if person $k$ is a friend of person $i$, and person $j$ is a friend of person $k$, i.e. person $j$ is a friend of a friend of $i$. Otherwise it is 0.

• when we sum 0/1 entries over all $k$ (all people on facebook), we count the 1s, i.e. $b_{ij}$ is the number of friends that person $i$ and person $j$ have in common.

• Facebook uses this construction: if $a_{ij} = 0$, but $b_{ij}$ is large, it will suggest $j$ as potential friend to $i$ (and vice versa) as potential new friends (because $i$ and $j$ have many common friends).

• We can go on with this: $C = AAA = A^3$:

$$c_{ij} = \sum_{k=1}^n \sum_{l=1}^n a_{ik}a_{kl}a_{lj}$$

is the number of connections between $i$ and $j$ with exactly 2 friends inbetween.

The 6 degrees of separation hypothesis states that for in a social network (not just on the Internet) in $A + A^2 + A^3 + A^4 + A^5 + A^6$ almost all entries are nonzero: “Everybody is connected to everybody in 6 steps or less.”

In (arXiv:1111.4503 [cs.SI]) this was tested for actual Facebook data:

• $A$ itself is very sparse: each row of $A$ has on average 140 ones, and 999,999,860 zeros.
but: 99.91% of Facebook users form a connected group.

99.6% of have a distance of 6 or less to each other.

the average distance between two Facebook users was 4.7

Example 6.6. Twitter Friendship in Facebook is symmetric: $a_{ij} = 1$ iff $a_{ji} = 1$. One can use matrices also for asymmetric relations, e.g. 'following' in Twitter.

Let $n$ be the number of Twitter users ($n \geq 500,000,000$ in Feb. 2012). Let $A \in \mathbb{R}^{n \times n}$ encode the 'follow' relation:

$$a_{ij} = \begin{cases} 1 & \text{if } j \text{ follows } i's \text{ tweet} \\ 0 & \text{otherwise.} \end{cases}$$

What is $B = A^2$, i.e. $b_{ij} = \sum_{k=1}^{n} a_{ik}a_{kj}$?

- $a_{ik}a_{kj} = 1$, iff user $k$ follows $i$ and $j$ follows $k$. $b_{ij}$ is the total number of people who are read by $j$, and who themselves read $i$. Again, Twitter uses entries with $a_{ij} = 0$ but $b_{ij} > 0$ to make suggestions.

Example 6.7. (Adjacency matrices). Facebook and Twitter are just special cases. Any graph $(V, E)$ can be encoded by a matrix. We order the vertices (arbitrarily), and write $A \in \mathbb{R}^{n \times n}$ for $|V| = n$ with $a_{ij} = 1$ if $(i, j) \in E$ and $a_{ij} = 0$ otherwise.

- Undirected graphs: if $(i, j) \in E$, then also $(j, i) \in E$, so $a_{ij} = a_{ji}$.

- Directed graphs: no such rule, we can have $a_{ij} \neq a_{ji}$.

6.2 Matrix Transpose

Sometimes, we might want to change the roles of rows and columns, e.g.

- EEG-matrix:
  1) each vector is a spatial field, we measure at multiple times, or
  2) each vector is a time series, we measured several of them at different locations.

- Adjacency graph: flipping rows and columns corresponds to inverting the arrows, i.e. inverting the "flow".
Definition 6.8. Transpose / Symmetric Matrices
Let $A \in \mathbb{R}^{m \times n}$ be matrix.

- The matrix $B \in \mathbb{R}^{m \times n}$ with $b_{ij} = a_{ji}$ is called transpose of $A$.
  We write $A^\top$ (sometimes $A^T$ or $A^t$) for $B$.
- A matrix $A \in \mathbb{R}^{m \times n}$ is called symmetric, if $A^\top = A$.

Observation. Only square matrices (i.e. $m = n$) can be symmetric, otherwise $A = A^\top$ is impossible.

Observation. Important special case: the transpose of a (column-)vector (size: $n \times 1$) is a row (size $1 \times n$). The latter is sometimes also called a row vector, but we avoid this.

Lemma 6.9. Properties of Matrix Transpose
For any matrices $A, B$ and any $\alpha \in \mathbb{R}$:

1) If $A + B$ is well defined, then $(A + B)^\top = A^\top + B^\top$.

2) For any $\alpha \in \mathbb{R}$, $(\alpha A)^\top = \alpha A^\top$

3) If $AB$ is well defined, then $(AB)^\top = B^\top A^\top$ (in particular: $B^\top A^\top$ is well defined).

Proof. 1) We check this directly:

$$[(A + B)^\top]_{ij} = [A + B]_{ji} = a_{ji} + b_{ji} = [A^\top]_{ij} + [B^\top]_{ij} = [A^\top + B^\top]_{ij}$$

2) We check this directly:

$$[\alpha A]^\top_{ij} = [\alpha A]_{ji} = \alpha a_{ji} = \alpha [A]_{ji} = \alpha [A^\top]_{ij}$$

3) First, we check that we can multiply $B^\top A^\top$ at all:

$A \in \mathbb{R}^{n \times m}$, $B \in \mathbb{R}^{k \times l}$. Because $AB$ exists, we know that $m = k$.

$B^\top \in \mathbb{R}^{l \times k}$, $A^\top \in \mathbb{R}^{m \times n}$. We can compute $B^\top A^\top$, because $k = m$.

Now we check the equality: let $C = AB$, $D = B^\top A^\top$.

$$[C]_{ij} = \sum_k a_{ik} b_{kj}, \quad \text{so} \quad [C^\top]_{ij} = [C]_{ji} = \sum_k a_{jk} b_{ki}.$$

$$[D]_{ij} = \sum_k [B^\top]_{ij} [A^\top]_{kj} = \sum_k b_{ji} a_{jk}$$

so indeed $C = D$. 

\[\square\]
Observation. Special case of 3): let $x, y \in \mathbb{R}^n$ be vectors. Then we can compute $x^\top y = \sum_{i=1}^n x_i y_i$, and this is a $1 \times 1$ matrix, i.e. a number. This allows us to write some expression we saw earlier more compactly:

- Let $f_1, \ldots, f_n$ be the payoff for $n$ outcomes of an experiment, stacked into a vector $f = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} \in \mathbb{R}^n$. Let $p_1, \ldots, p_n$ be the probabilities of the $n$ possible outcomes, stacked into a vector $p = \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix} \in \mathbb{R}^n$. The expected payoff, $\mathbb{E}_p\{f\} = \sum_i p_i f_i = p^\top f$.

- Let $A = (a_{ij})$ be the Facebook friend-matrix. The average number of friends per person is given by

$$\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n a_{ij} = \frac{1}{n}1^\top A 1 \quad (6.6)$$

Caveat: for $x, y \in \mathbb{R}^n$, $x^\top y$ is single number, but $xy^\top$ is a $n \times n$ matrix!

Example 6.10.

1) A matrix, $A$, of pairwise distances, $a_{ij} = d(x^{(i)}, x^{(j)})$, is symmetric.

2) The identity matrix $\text{Id}_{n \times n}$ is symmetric, as is $\alpha \text{Id}_{n \times n}$ for any $\alpha \in \mathbb{R}$.

3) For any square matrix $A \in \mathbb{R}^{n \times n}$, the matrix $A + A^\top$ is symmetric.

4) For any $A \in \mathbb{R}^{n \times m}$, $AA^T \in \mathbb{R}^{n \times n}$ and $A^T A \in \mathbb{R}^{m \times m}$ are symmetric.

5) 2D-rotation matrices are in general not symmetric (exceptions: integer multiples of 180 degrees: 0, 180, 360, \ldots)

6) The Facebook friendship matrix is symmetric.

7) The Twitter following matrix is square, but not symmetric.