Lecture 2: A Geometric View of Linear Algebra.

* High dimensional vectors: $v \in \mathbb{R}^d$ (d: a large number)
* In ML, our data points are represented often as such vectors. Each coordinate of v corresponds to a feature:

Example: Flu detection

v =

0	1	99	....	(others)
Fever or Not	Sore throat or Not	Body temperature		

* Vector Operations and their geometry is important to understand to be able to design ML algorithms.

1. Vector Addition

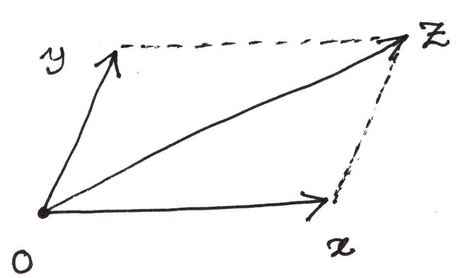
Let $x = [x_1, \dots, x_d]$

$y = [y_1, \dots, y_d]$

Then: $z = x + y = [x_1 + y_1, \dots, x_d + y_d]$, let $z' = x - y$

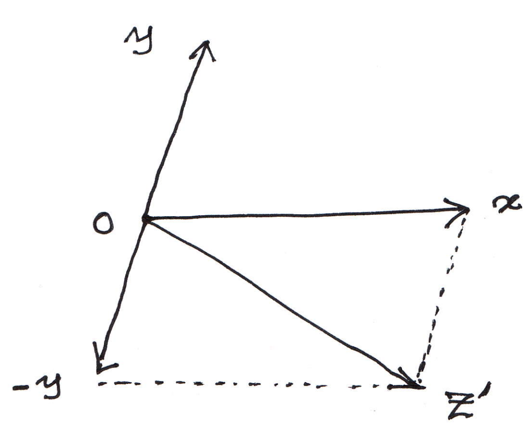
Geometrically: O = origin.

$z = x + y$



Parallelogram Law.

$z' = x - y$

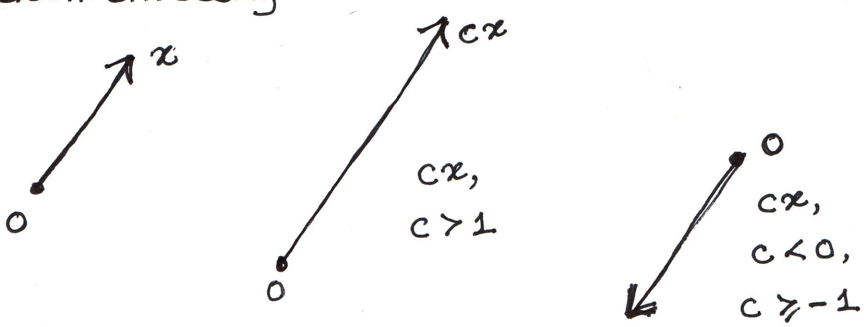


2. Scaling Vectors Up or Down.

Let $x = [x_1, .., x_d]$, c = a scalar.

let $z = cx$

Geometrically :



O = origin

3. Inner Products (or, Dot Products) and Norms.

* If $x = [x_1, .., x_d]$, then the norm of x is defined as:

$$\|x\| = \sqrt{\sum_{i=1}^{d} x_i^2}$$

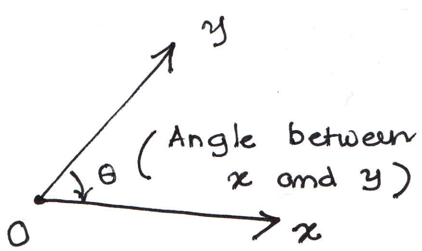
Geometrically, $\|x\|$ is the length of x

* For $x = [x_1, .., x_d]$, $y = [y_1, .., y_d]$, the inner product or dot product of x and y is defined as:

$$\langle x, y \rangle = \sum_{i=1}^{d} x_i y_i$$

- Geometrically, dot products are related to angles.

$$\cos(\text{Angle b/w } x \text{ and } y) = \frac{\langle x, y \rangle}{\|x\| \cdot \|y\|}$$



Some Properties and Facts:

1. $||x||^2 = \langle x, x \rangle$.

Proof: $$||x||^2 = \sum_{i=1}^{d} x_i^2 \quad \text{(By definition)}$$

$$\langle x, x \rangle = \sum_{i=1}^{d} x_i \cdot x_i = \sum_{i=1}^{d} x_i^2 \quad \text{(By definition)}$$

Verify: $\cos\theta = \cos 0^\circ = 1$ (in this case)

$$\frac{\langle x, x \rangle}{||x|| \cdot ||x||} = 1 \quad \text{also.}$$

2. Properties of the dot product:

- $\langle x+y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$
- $\langle x, y \rangle = \langle y, x \rangle$
- For scalar c, $\langle cx, y \rangle = c \langle x, y \rangle$

Proofs follow easily from the definitions.

3. The Euclidean distance between two vectors x and y is defined as:

$$\sqrt{\sum_{i=1}^{d} (x_i - y_i)^2} = ||x - y|| \quad \text{(observe)}$$

4. **Fact:** $||x-y||^2 = ||x||^2 + ||y||^2 - 2\langle x, y \rangle$

Proof: $$||x-y||^2 = \langle x-y, x-y \rangle \quad \text{(From (1))}$$

$$= \langle x, x \rangle + \langle y, y \rangle - 2\langle x, y \rangle$$

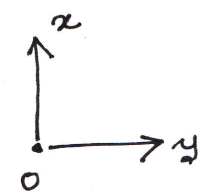
(From properties of dot products)

$$= ||x||^2 + ||y||^2 - 2\langle x, y \rangle$$

5. When x and y are orthogonal,

$$\langle x, y \rangle = 0$$

[Recall, $\cos\theta = \dfrac{\langle x, y\rangle}{\|x\|\,\|y\|} = \cos 90^\circ = 0.$]



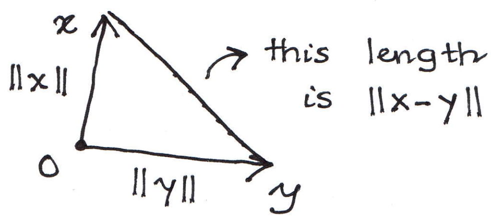
6. If x and y are orthogonal,

$$\|x-y\|^2 = \|x\|^2 + \|y\|^2$$

Proof: From Fact (4) and Fact (5).

Geometrically, this statement is Pythagora's theorem

(so this is a one-line proof)



7. Cauchy Schwartz Inequality:

For any vectors x and y,

$$\langle x, y\rangle \le \|x\| \cdot \|y\|$$

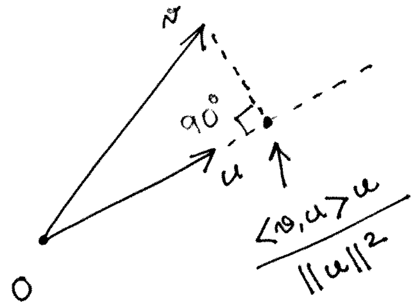
Proof: For any vectors x and y, $\cos(\theta) \le 1$, where θ is the angle between them.

Also: $\cos\theta = \dfrac{\langle x, y\rangle}{\|x\|\,\|y\|}$

So, we have: $\langle x, y\rangle \le \|x\|\,\|y\|$

(One line proof.)

Projections and Dot Products:



Look at plane containig u, v.

Draw a perpendicular from v to u.

Projection of v along u is the vector $\dfrac{\langle v, u\rangle u}{\|u\|^2} = a$

This is also called "component" of v along u.

Component of v perpendicular to u is $v - \dfrac{\langle v, u\rangle u}{\|u\|^2} = b$.

<u>Fact</u>: a, b are orthogonal.

<u>Proof</u>: $\langle a, b\rangle = \left\langle \dfrac{\langle v, u\rangle u}{\|u\|^2}, v - \dfrac{\langle v, u\rangle u}{\|u\|^2} \right\rangle$

$$= \frac{\langle v, u\rangle \langle u, v\rangle}{\|u\|^2} - \frac{\langle v, u\rangle^2}{\|u\|^4} \cdot \underbrace{\langle u, u\rangle}_{\text{this is } \|u\|^2}$$

$$= \frac{\langle v, u\rangle^2}{\|u\|^2} - \frac{\langle v, u\rangle^2}{\|u\|^2} = 0$$

<u>Observe</u>: If u is a unit vector, $(\|u\| = 1)$ then:

- Length of projection of v along u is $\langle v, u\rangle$
- Component of v orthogonal to u is $v - \langle v, u\rangle u$.

Matrices and Matrix Multiplication:

A $n \times d$ matrix M is written as:

$$M = \begin{bmatrix} M_{11} & M_{12} & \cdots & M_{1d} \\ M_{21} & M_{22} & \cdots & M_{2d} \\ \vdots & & & \\ M_{n1} & M_{n2} & \cdots & M_{nd} \end{bmatrix}$$

n rows, d columns

M_{ij} : entry at row i, column j

M^T : Transpose of M: $d \times n$ matrix : $M^T_{ij} = M_{ji}$

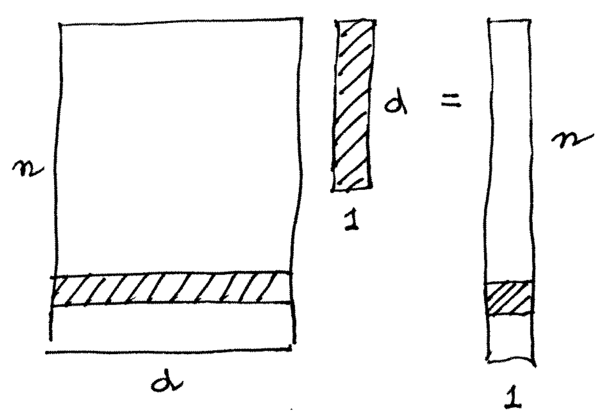
Matrix Vector Multiplication:

M: $n \times d$ matrix

x: $d \times 1$ vector

Mv: $n \times 1$ vector z s.t.

$$z_i = \sum_{j=1}^{d} M_{ij} x_j$$



Observe: $z_i = \langle M_i, x \rangle$

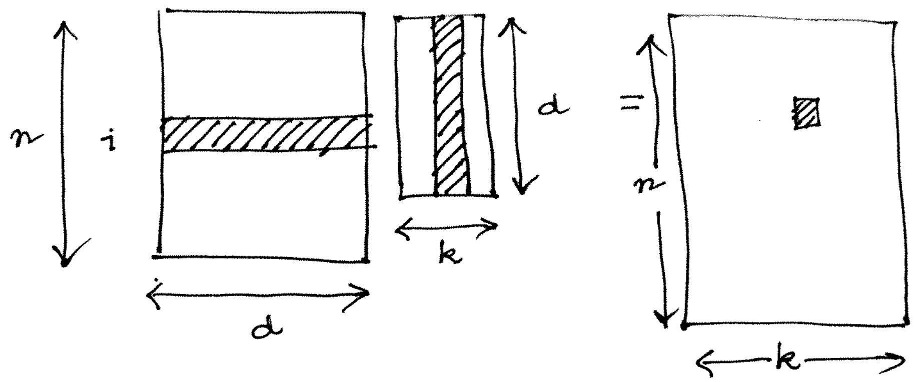
z_i : Length of the projection of M_i onto x when x is a unit vector.

Matrix - Matrix Multiplication:

$M_{n \times d}$ matrix; $N_{d \times k}$

MN : $n \times k$ matrix P such that

$$P_{ij} = \sum_{l=1}^{d} M_{il} N_{lj}$$



Observe:

$$P_{ij} = \langle M_i, n_j \rangle$$

$M_i \downarrow$ Row i of M

$n_j \rightarrow$ Column j of N

P_{ij} : length of projection of M_i on n_j when n_j is a unit vector.

Identity and Diagonal Matrices:

- M is diagonal if $M_{ij} = 0$ for $i \neq j$.
- M is identity if M is ~~ad~~ diagonal and $M_{ii} = 1$, for all i. and M is $d \times d$.

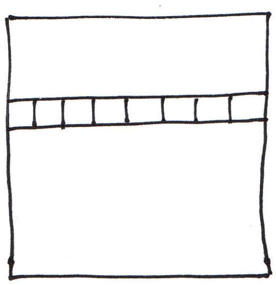
Data Matrix:

$n \times d$ matrix. A

n data points: $x^1, .., x^n$

Each row of A is a data point.

[This is a special matrix we will use often in this class.]

A =  $= x^i$

* Let u be an unit $d \times 1$ vector. How can we enterpret Au?
* Let $U = [u_1, .., u_k]$ be a $d \times k$ matrix whose ~~as~~ i-th column u_i is an unit vector for all i. How to uinterpret AU?

Some Identities:

1. For a $d \times 1$ vector x, $x^T x = \langle x, x \rangle = \|x\|^2$

$$x^T x = \sum_{i=1}^{d} x_i^2 = \|x\|^2 = \langle x, x \rangle \quad \text{(By definition)}$$

2. For $d \times 1$ vectors x and y, $x^T y = \langle x, y \rangle$

3. Let A be a $d \times d$ matrix and x be a $d \times 1$ vector.

Then:
$$x^T A x = \sum_{i=1}^{d} \sum_{j=1}^{d} A_{ij} x_i x_j$$

Proof: $z = Ax$ is $d \times 1$ vector s.t.

$$z_i = \sum_{j=1}^{d} A_{ij} x_j$$

Now,
$$x^T A x = x^T z = \sum_{i=1}^{d} x_i z_i = \sum_{i=1}^{d} \sum_{j=1}^{d} A_{ij} x_i x_j$$

Positive Semi-Definite (PSD) Matrix:

A matrix $A_{d\times d}$ is PSD if for all $d\times 1$ vectors x,

$$x^T A x \geq 0$$

Examples:

1. $I_{d\times d} = \begin{pmatrix} 1 & & 0 \\ & 1 & \\ 0 & & \ddots \; 1 \end{pmatrix}$ is PSD.

For any x, $x^T I x = \sum_{i,j} x_i x_j I_{ij} = \sum_i x_i^2 \geq 0$.

2. $\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$ is PSD.

For any $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $x^T A x = 2x_1^2 + x_2^2 - 2x_1 x_2$

$$= (x_1 - x_2)^2 + x_2^2 \geq 0$$

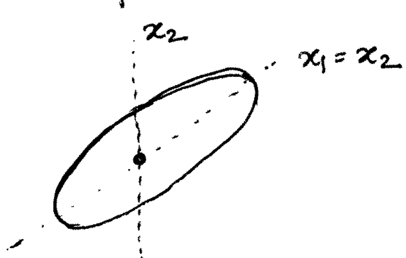
3. $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is NOT PSD. Why?

Take $x = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$, $x^T A x = 2(-1)1 = -2 < 0$

Geometry: Symmetric PSD matrices can be represented by ellipsoids in high dimensions.

eg: $I_{d\times d}$ represents a d-dim sphere with eqn: $\sum_{i=1}^{d} x_i^2 = 1$

$\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$ represents the ellipse $(x_1 - x_2)^2 + x_2^2 = 1$.



Vector Subspaces:

Let $v_1, .., v_k$ be k vectors. The subspace spanned by them is the set of all vectors of the form:

$$\sum_{i=1}^{k} a_i v_i \qquad (a_i = \text{a scalar, for all } i)$$

Geometrically:

- v_1, $v_2 = 2v_1$.

  Subspace spanned: any linear combination of v_1. (a line)

- v_1, v_2 (where v_2 is linearly independent of v_1)

  Subspace spanned: plane containing v_1 and v_2

Dimension of a subspace is the maximum # of linearly independent vectors in the subspace.

- Dimension of a subspace spanned by $v_1, .., v_k$ is $\leq k$.

Important Relationship:

Let $x^1, .., x^n$ be n data vectors.

Dimension of subspace spanned by $x^1, .., x^n$ = Rank of the data matrix A.

If your data is inherently low dimensional, then data matrix has low rank.