Recitation 4

Alex Dong

CDS, NYU

Fall 2020

Norms

- Norms measure distances!
- Think about all the "natural" properties of distance that make sense.
    - distance $= 0$ means at the same point
    - distance is always non-negative
    - distance follows triangle inequality (at least in Euclidean space)

Norms

Shorthand way to remember what the properties do.

Definition (Norm)

A norm $\| \cdot \|$ on V verifies the following points:

1. *Triangular inequality*: $\|u + v\| \leq \|u\| + \|v\|$ "Euclidean space"
2. *Homogeneity*: $\|\alpha v\| = |\alpha| \times \|v\|$ "farther actually means farther"
3. *Positive definiteness*: if $\|v\| = 0 \implies v = 0$. "Non-negative"

Inner Products

Definition (Inner product)

Let V be a vector space. An inner product on V is a function $\langle \cdot, \cdot \rangle$ from $V \times V$ to $\mathbb{R}$ that verifies the following points:

1. *Symmetry*: $\langle u, v \rangle = \langle v, u \rangle$ for all $u, v \in V$.
2. *Linearity*: $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$ and $\langle \alpha v, w \rangle = \alpha \langle v, w \rangle$ for all $u, v, w \in V$ and $\alpha \in \mathbb{R}$.
3. *Positive definiteness*: $\langle v, v \rangle \geq 0$ with equality if and only if $v = 0$.

- Definition of inner product does not reveal it's purpose.
- **In this class, we always use the Euclidean inner product.**
    - $\langle u, v \rangle = u^T v$
- (!!) Inner products are (indirectly) used for a notion of angles.
- $cos(\theta) = \frac{\langle u,v \rangle}{\|u\| \|v\|}$

Inner Products in Machine Learning (&)

- Inner products can be used as a measure of similarity
- Kernel Tricks (&) - Increase Data Complexity
  - Sometimes you have to calculate $x_{old}^T x_{new}$, equivalently $\langle x_{old}, x_{new} \rangle$
  - You can replace the inner product with a inner product in a higher dimensional space
  - Instead of calculating $\langle x_{old}, x_{new} \rangle$, define a function K and calculate $\langle K(x_{old}), K(x_{new}) \rangle$
  - If you pick "the right" higher dimensional space, your data can be a lot easier to work with

[0] (&) denotes extra material not covered in this course

Questions 1: Norms and Inner Products

1. Which of the following functions are inner products for $x, y \in \mathbb{R}^3$?
   i. $f(x, y) = x_1 y_2 + x_2 y_3 + x_3 y_1$
   ii. $f(x, y) = x_1^2 y_1^2 + x_2^2 y_2^2 + x_1^2 y_1^2$
   iii. $f(x, y) = x_1 y_1 + x_3 y_3$
2. For $A \in \mathbb{R}^{m \times n}$ and $x \in \mathbb{R}^n$, prove that

$$\|Ax\| \leq \|x\| \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} A_{i,j}^2}$$

Solutions 1: Norms and Inner Products

1. Which of the following functions are inner products for $x, y \in \mathbb{R}^3$?

Solution

i. $f(x, y) = x_1y_2 + x_2y_3 + x_3y_1$ *False*
Consider $u = [1, 0, 0]^T$ *and* $v = [0, 1, 0]^T$.
$\langle u, v\rangle = 1$, *but* $\langle v, u\rangle = 0$. *(Not symmetric)*

ii. $f(x, y) = x_1^2y_1^2 + x_2^2y_2^2 + x_3^2y_3^2$ *False*
Consider $v = [1, 0, 0]^T$.
$\langle 2v, v\rangle = 4$, *but* $2\langle v, v\rangle = 2$. *(Not linear)*

iii. $f(x, y) = x_1y_1 + x_3y_3$ *False*
Consider $v = [0, 1, 0]^T$.
$\langle v, v\rangle = 0$, *but* $v \neq 0$. *(Not positive definite)*

2. For $A \in \mathbb{R}^{m\times n}$ and $x \in \mathbb{R}^n$, prove that

$$\|Ax\| \leq \|x\| \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} A_{i,j}^2}$$

Solution

Let $A = \begin{bmatrix} - & \mathbf{a_1}^T & - \\ \vdots & \vdots & \vdots \\ - & \mathbf{a_m}^T & - \end{bmatrix}$ *and* $x = \begin{bmatrix} | \\ x \\ | \end{bmatrix}$. *Observe that* $Ax = \begin{bmatrix} \langle \mathbf{a_1}, x\rangle \\ \vdots \\ \langle \mathbf{a_m}, x\rangle \end{bmatrix}$.

Now,

$\|Ax\|^2 = \sum_{i=1}^{m} |\langle \mathbf{a_i}, x\rangle^2|$ *by definition of norm*

$\|Ax\|^2 \leq \sum_{i=1}^{m} \|\mathbf{a_i}\|^2 \|x\|^2$ *by Cauchy-Schwarz*

$\|Ax\| \leq (\sum_{i=1}^{m} \|\mathbf{a_i}\|^2 \|x\|^2)^{.5}$

$\|Ax\| \leq \|x\| (\sum_{i=1}^{m} \|\mathbf{a_i}\|^2)^{.5}$

$\|Ax\| \leq \|x\| (\sum_{i=1}^{m} \sum_{j=1}^{n} A_{i,j})^{.5}$ *by definition of* $\mathbf{a_i}$

Orthogonality

- Angles can be used as a measure of similarity
- Vectors u, v are orthogonal if and only if $\langle u, v \rangle = 0$
- Vectors are orthogonal $\implies$ vectors are as dissimilar as possible
- Orthogonal coordinate systems are nice because we can view each coordinate "independently" (we will prove later).
- Gram-Schmidt Process (Lec 5) allows us to change any basis into an orthonormal basis.

Orthogonal Projections

- Projections form an important part of linear algebra.
  - We can view the action of a matrix and how it affects a certain subspace
  - We can simplify our data by picking the subspace "closest" to the data (PCA, Lec 7)
  - We can find the best-fit line/plane/subspace (Linear regression, Lec 9)
- *Orthogonal* projections are a special kind of projection
  - They preserve the original vector components (in the orthogonal basis)

Questions: Orthogonality

1. Let $v_1, ..., v_k$ be a list of orthogonal vectors. Show that $v_1, ..., v_k$ are linearly independent.
2. Let U be the subspace of $\mathbb{R}^n$ with orthonormal basis $u_1, ..., u_k$.
   i. Prove that the orthogonal projection of $v \in \mathbb{R}^n$ onto U can be expressed as $P_U = \sum_{i=0}^{k} \langle v, u_i \rangle u_i$. (Use the fact that the orthonormal basis for a subspace of $\mathbb{R}$ can be extended to obtain an orthonormal basis for $\mathbb{R}$)
   ii. Prove that $P_U(v) \leq \|v\|$
   iii. Prove that $v - P_U(v)$ is orthogonal to $P_U(v)$

Solutions: Orthogonality

Solution

1. *Let $v_1, ..., v_k$ be a list of non-zero orthogonal vectors. Show that $v_1, ..., v_k$ are linearly independent.*

Let $\alpha_1, ..., \alpha_k \in \mathbb{R}$ *s.t* $\sum_{i=1}^{k} \alpha_i v_i = \vec{0}$.
Consider $\langle \sum_{i=1}^{k} \alpha_i v_i, \sum_{j=1}^{k} \alpha_j v_j \rangle$.

$$
\begin{aligned}
0 = & \ \langle \vec{0}, \vec{0} \rangle \\
= & \ \langle \sum_{i=1}^{k} \alpha_i v_i, \sum_{j=1}^{k} \alpha_j v_j \rangle \\
= & \sum_{i=1}^{k} \alpha_i^2 \langle v_i, v_i \rangle, \sum_{i \neq j} \alpha_i \alpha_j \langle v_i, v_j \rangle \\
0 = & \sum_{i=1}^{k} \alpha_i^2 \qquad \textit{by orthonormality of } v_i, v_j
\end{aligned}
$$

So $\alpha_1, ..., \alpha_k = 0$.

Solutions: Orthogonality

Solution

Let U be the subspace of $\mathbb{R}^n$ with orthonormal basis $u_1, ..., u_k$.

2i. *Prove that the orthogonal projection of $v \in \mathbb{R}^n$ onto U can be expressed as $P_U(v) = \sum_{i=0}^{k} \langle v, u_i \rangle u_i$.*

Let $u_{k+1}, ..., u_n$ be an orthonormal basis extension for $u_1, ..., u_k$.

Then $u_1, ..., u_k, u_{k+1}, ..., u_n$ form an orthonormal basis for $\mathbb{R}^n$

Now, let $v = \sum_{i=1}^{n} \alpha_i u_i$ where $\alpha_i = \langle v, u_i \rangle$ and let $x \in U$, where $x = \sum_{j=1}^{k} \beta_i u_i$.

We want to find $\arg\min_{x \in U} \|v - x\|$.

$$\begin{aligned}
\|v - x\| &= \| \sum_{i=1}^{n} \alpha_i u_i - \sum_{j=1}^{k} \beta_i u_i \| \\
&= \| \sum_{j=1}^{k} (\alpha_i - \beta_i) u_i - \sum_{i=k+1}^{n} \alpha_i u_i \| \\
&= \sqrt{\sum_{j=1}^{k} (\alpha_i - \beta_i)^2 + \sum_{i=k+1}^{n} \alpha_i^2} \qquad \textit{by orthonormality}
\end{aligned}$$

$\|v - x\|$ is minimized when $\alpha_i = \beta_i \quad \forall i \in \{1, ..., k\}$

This implies that $\beta_i = \langle v, u_i \rangle$.

So $P_U(v) = argmin_{x \in U} \|v - x\| = \sum_{i=0}^{k} \langle v, u_i \rangle u_i$.

Solutions: Orthogonality

Solution

Let U be the subspace of $\mathbb{R}^n$ with orthonormal basis $u_1, ..., u_k$.

2ii. Prove that $P_U(v) \leq \|v\|$

$P_U(v) = \sum_{i=1}^k \langle v, u_i \rangle u_i$ *from 2i*

$$\begin{aligned}
\|P_U(v)\|^2 &= \|\sum_{i=1}^k \langle v, u_i \rangle u_i\|^2 \\
&= \sum_{i=1}^k \|\langle v, u_i \rangle u_i\|^2 && \textit{by Pythagorean Theorem} \\
&\leq \sum_{i=1}^n \|\langle v, u_i \rangle u_i\|^2 && \textit{add extra components} \\
&= \|\sum_{i=1}^n \langle v, u_i \rangle u_i\|^2 && \textit{Pythagorean Theorem} \\
&= \|v\|^2
\end{aligned}$$

So $P_U(v) \leq \|v\|$

Solutions: Orthogonality

Solution

Let U be the subspace of $\mathbb{R}^n$ with orthonormal basis $u_1, ..., u_k$.

2iii. Prove that $v - P_U(v)$ is orthogonal to $P_U(v)$

$P_U(v) = \sum_{i=1}^{k} \langle v, u_i \rangle u_i$ *from 2i*

$v = \sum_{i=0}^{n} \langle v, u_i \rangle u_i$ *since $u_1, ..., u_n$ is a orthonormal basis.*

$$\begin{aligned} v - P_U(v) &= \sum_{i=1}^{n} \langle v, u_i \rangle u_i - \sum_{i=1}^{k} \langle v, u_i \rangle u_i \\ &= \sum_{i=k+1}^{n} \langle v, u_i \rangle u_i \end{aligned}$$

$$\begin{aligned} \langle v - P_U(v), v \rangle &= \langle (\sum_{i=k+1}^{n} \langle v, u_i \rangle u_i), (\sum_{i=1}^{k} \langle v, u_i \rangle u_i) \rangle \\ &= 0 \qquad u_i \text{ are pairwise orthogonal.} \end{aligned}$$

Questions: Orthogonal Complements

Let S, U be subspaces of a vector space V.
Prove the following statement:

1. $S \subset U \implies S^\perp \supset U^\perp$

Let $A \in \mathbb{R}^{n \times m}$. Assume the Euclidean inner product.

2. (!) Prove that $Im(A^T) = Ker(A)^\perp$.

   (Hint: $\implies$ is easy. Use (1) for $\impliedby$)

Solutions: Orthogonal Complements

1. $S \subset U \implies S^{\perp} \supset U^{\perp}$

Solution

Let $x \in U^{\perp}$*, and* $z \in S$*.*
Since $z \in S$ *and* $S \subset U$*, then* $z \in U$*.*
Now, since $x \in U^{\perp}$ *and* $z \in U$*, then* $\langle x, z \rangle = 0$*.*
So $x \in S^{\perp}$*.*

Solutions: Orthogonal Complements

2. Prove that $Im(A^T) = Ker(A)^{\perp}$.

Solution

$\Longrightarrow$

Let $x \in \mathrm{Im}(A^T)$. *Then* $\exists y$ *s.t* $x = A^T y$. *We show* $x \in Ker(A)^{\perp}$.
Let $v \in Ker(A)$. *Then* $Av = 0$.
Consider $\langle x, v\rangle$.

$$\langle x, v\rangle = x^T v = y^T Av = \langle y, Av\rangle = \langle y, 0\rangle = 0 \text{ Then } x \in Ker(A)^{\perp}.$$

$\Longleftarrow$.

We use 1. to show $Im(A^T)^{\perp} \subset Ker(A)$ *instead.*
Let $x \in Im(A^T)^{\perp}$.
Consider Ax. *We show* $\langle x, A^T y\rangle = 0$ *for all* $y \in \mathbb{R}^n$.
Since $x \in Im(A^T)^{\perp}$, *then* $\forall y \in \mathrm{Im}(A^T)$, $\langle x, y\rangle = x^T y = 0$.
Consider $\|Ax\|$.

$$\|Ax\| = x^T A^T Ax = x(A^T Ax).$$

Since $A^T Ax \in Im(A^T)$, *then* $\|Ax\| = 0$, *so* $Ax = 0$.
Now, by 1, we can conclude that $Ker(A)^{\perp} \subset Im(A^T)$.

Appendix starts after here

Idempotence

Lets take a step back.

- P_S is an *orthogonal* projection $\iff P_S = VV^T$
  - V has orthonormal columns that form a basis for S.
- There is a more general definition of a projection - known as *idempotence*.

Definition (Idempotence)

An matrix P is idempotent when $P^2 = P$.

An idempotent matrix is also called a *projection* or *projection matrix*.

Questions: Orthogonal Projections vs Idempotence

Definition (Idempotence)

An matrix P is idempotent when $P^2 = P$.

1. Show that $X(X^T X)^{-1} X^T$ is idempotent.
2. Show that all orthogonal projections are idempotent.
3. Give an example of an idempotent matrix that is not an orthogonal projection.
   (Hint: Show that your matrix does not minimize the distance to subspace it projects onto.)

Solutions: Orthogonal Projections vs Idempotence

Solution

1. *Show that $X(X^TX)^{-1}X^T$ is idempotent.*

$$\begin{aligned} P^2 &= (X(X^TX)^{-1}X^T)(X(X^TX)^{-1}X^T) \\ &= X(X^TX)^{-1}(X^TX)(X^TX)^{-1}X^T \\ &= X(X^TX)^{-1}X^T \end{aligned}$$

2. *Show that all orthogonal projections are idempotent.*

   Let P be an orthogonal projection.
   Recall that all orthogonal projections take the form VV^T, where $V \in \mathbb{R}^{n\times k}$ has orthonormal columns.
   Note that $V^TV = I_k$, the identity matrix in $\mathbb{R}^{k\times k}$.
   Then $P^2 = (VV^T)(VV^T) = V(V^TV)V^T = VI_kV^T = VV^T = P$

Solutions: Orthogonal Projections vs Idempotence

Solution

3. *Give an example of an idempotent matrix that is not an orthogonal projection.*

Consider the matrix $A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$

It's easy to see $A^2 = A$, *and* $Im(A) = \{ \begin{bmatrix} x \\ 0 \end{bmatrix} \mid x \in \mathbb{R} \}$

Consider the vector $v = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$

The closest vector in Im(A) is $v_{Im(A)} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$, *but* $Av = \begin{bmatrix} 4 \\ 0 \end{bmatrix}$

Note: Rigorously speaking, we need to prove that $v_{Im(A)} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$ *is the closest vector in* $Im(A)$. *We can do this by constructing an orthogonal projection onto* $Im(A)$, *which is found by setting* $V = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, *and calculating* $VV^T = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$