INNER PRODUCT AND ITS APPLICATIONS IN DATA SCIENCE

Dinh Huu Nguyen, 04/05/2018

Abstract: an exposition on inner product, that helps to answer such questions as

- whether Mahalanobis distance satisfies triangle inequality

$$d_M(u, w) \leq d_M(u, v) + d_M(v, w)$$

- how averaging may reduce overfitting

Contents

1. Vector Spaces

We begin with the real line without any structure

$$\mathbb{R} = \{\text{all real numbers}\}$$

Then we equip it with addition

$$(\mathbb{R}, +)$$

Then we equip it with multiplication by real numbers that is compatible with addition

$$(\mathbb{R}, +, \cdot)$$

Definition 1.1. A vector space is a set V together with a binary operation $+$ and a scalar multiplication $\cdot$ that satisfy the following axioms

1. (closure under addition) $u + v \in V$ for any $u, v \in V$.
2. (associativity of addition) $(u + v) + w = u + (v + w)$ for all $u, v, w \in V$.
3. (commutativity of addition) $u + v = v + u$.
4. (identity element under addition) there exists an element $0 \in V$ such that $0 + v = v + 0 = v$ for all $v \in V$.
5. (inverse element under addition) there exists an element $-v$ such that $v + (-v) = -v + v = 0$ for all $v \in V$.
6. (closure under scalar multiplication) $a \cdot v \in V$ for any $a \in \mathbb{R}$ and $v \in V$.
7. (distributivity of scalar multiplication with respect to vector addition) $a \cdot (u + v) = a \cdot u + a \cdot v$.
8. (distributivity of scalar multiplication with respect to field addition) $(a + b) \cdot v = a \cdot v + b \cdot v$.
9. (compatibility of scalar multiplication with field multiplication) $(ab) \cdot v = a \cdot (b \cdot v)$.
10. (identity element of scalar multiplication) $1 \cdot v = v$ for $1 \in \mathbb{R}$ and any $v \in V$.

Example 1.2. A single point $V = \{*\}$ together with binary operation

$$* + * = *$$

and scalar multiplication

$$a \cdot * = *$$

for any $a \in \mathbb{R}$ is a vector space. We often call it the 0 vector space and denote it as $V = \{0\}$.

Example 1.3. The Euclidean space $\mathbb{R}^m$ together with binary operation

$$(u_1, \ldots, u_m) + (v_1, \ldots, v_m) = (u_1 + v_1, \ldots, u_m + v_m)$$

and scalar multiplication

$$a \cdot (u_1, \ldots, u_m) = (au_1, \ldots, au_m)$$

for any $(u_1, \ldots, u_m), (v_1, \ldots, v_m) \in \mathbb{R}^m$ and $a \in \mathbb{R}$ is a vector space.

Example 1.4. The set $RV((\Omega, \mathcal{F}, P), (\mathbb{R}, \mathcal{B}(\mathbb{R})))$ of all random variables from $(\Omega, \mathcal{F}, P)$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ form a vector space.

Example 1.5. The subset $RVZF((\Omega, \mathcal{F}, P), (\mathbb{R}, \mathcal{B}(\mathbb{R})))$ of all random variables with zero mean and finite variance form a subspace of $RV((\Omega, \mathcal{F}, P), (\mathbb{R}, \mathcal{B}(\mathbb{R})))$.

2. Inner Product

Definition 2.1. An inner product on a vector space V is any map

$$V \times V \xrightarrow{\langle -,- \rangle} \mathbb{R}$$
$$(u, v) \mapsto \langle u, v \rangle$$

that satisfies the following axioms

1. (symmetry) $\langle u, v \rangle = \langle v, u \rangle$.
2. (linearity) $\langle a \cdot u + b \cdot v, w \rangle = a\langle u, w \rangle + b\langle v, w \rangle$.
3. (positive definiteness) $\langle v, v \rangle \geq 0$ for all $v \in V$, with equality iff $v = 0$.

A vector space V equipped with an inner product is called an inner product space.

Example 2.2. The usual inner product on $\mathbb{R}^m$ is

$$\mathbb{R}^m \times \mathbb{R}^m \xrightarrow{\langle -,- \rangle_I} \mathbb{R}$$
$$((u_1, \dots, u_m), (v_1, \dots, v_m)) \mapsto u_1 v_1 + \cdots + u_m v_m$$

This inner product can also be written as

$$\langle (u_1, \dots, u_m), (v_1, \dots, v_m) \rangle_I = \begin{pmatrix} u_1 & \cdots & u_m \end{pmatrix} \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_m \end{pmatrix}$$

Example 2.3. In fact each inner product is defined by a positive definite matrix A and vice versa.

$$\langle u, v \rangle_A = u^t A v$$

Example 2.4. If we use $A = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}$ then

$$\begin{aligned} \langle (u_1, u_2), (v_1, v_2) \rangle_A &= \begin{pmatrix} u_1 & u_2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \\ &= u_1 v_1 + 4 u_2 v_2 \end{aligned}$$

3. Norm

We can define different norms to measure how large a vector $v \in V$ is, and they all must meet three criteria.

Definition 3.1. A norm on a vector space V is a map

$$V \xrightarrow{||-||} \mathbb{R}$$
$$v \mapsto ||v||$$

that satisfies

1. (positive definiteness) $||v|| \geq 0$, with equality iff $v = 0$.
2. (homogeneity) $||a \cdot v|| = |a|\,||v||$.
3. (triangle inequality) $||v + w|| \leq ||v|| + ||w||$.

A vector space V equipped with a norm is called a normed space.

Example 3.2. The 1-norm on $\mathbb{R}^m$ is

$$\begin{aligned} \mathbb{R}^m &\xrightarrow{||-||_1} \mathbb{R} \\ (v_1, \ldots, v_m) &\mapsto |v_1| + \cdots + |v_m| \end{aligned}$$

Example 3.3. The 2-norm on $\mathbb{R}^m$ is

$$\begin{aligned} \mathbb{R}^m &\xrightarrow{||-||_2} \mathbb{R} \\ (v_1, \ldots, v_m) &\mapsto \sqrt{v_1^2 + \cdots + v_m^2} \end{aligned}$$

Example 3.4. The usual inner product in example 2.2 induces the 2-norm

$$||v||_2 = \sqrt{\langle v, v\rangle_I}$$

Example 3.5. The inner product in example 2.3 induces the norm

$$||v||_A = \sqrt{\langle v, v\rangle_A}$$

Example 3.6. In fact every inner product induces a norm

$$||v|| = \sqrt{\langle v, v\rangle}$$

4. Distance

A norm induces distance.

Definition 4.1. We define the distance between two vectors u, v in a normed space $(V, || - ||)$ as

$$d(u, v) = ||u - v||$$

Example 4.2. If we use the usual inner product $\langle -,-\rangle_I$ and its induced 2-norm $||-||_2$ in example 3.4 on $\mathbb{R}^2$ then

$$\begin{aligned} d_2((0,0),(0,1)) &= ||(0,0)-(0,1)||_2 \\ &= ||(0,1)||_2 \\ &= 1 \end{aligned}$$

Example 4.3. If we use the inner product $\langle -,-\rangle_A$ and its induced norm $||-||_A$ in example **??** on $\mathbb{R}^2$ then

$$\begin{aligned} d_A((0,0),(0,1)) &= ||(0,0)-(0,1)||_A \\ &= ||(0,1)||_A \\ &= \sqrt{\langle (0,1),(0,1)\rangle_A} \\ &= \sqrt{\begin{pmatrix} 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix}} \\ &= 2 \end{aligned}$$

Example 4.4. (Mahalanobis distance) Suppose your dataset $X_{n\times m}$ of n samples and m features has covariance matrix M^{-1} with inverse M

Then we can use it to define Mahalanobis inner product

$$\mathbb{R}^m \times \mathbb{R}^m \xrightarrow{\langle -,- \rangle_M} \mathbb{R}^m$$
$$(u, v) \mapsto u^t M v$$

This Mahalanobis inner product induces the Mahalanobis norm $||-||_M$ and the Mahalanobis distance d_M.

Example 4.5. (Mahalanobis distance satisfies triangle inequality) We compare $d_M(u, w)$ and $d_M(u, v) + d_M(v, w)$

$$\begin{aligned} d_M(u, w) &= ||u - w||_M \\ &= ||u - v + v - w||_M \\ &\leq ||u - v||_M + ||v - w||_M \text{ (triangle inequality)} \\ &= d_M(u, v) + d_M(v, w) \end{aligned}$$

5. Correlation

An inner product and its induced norm induce correlation coefficient.

Definition 5.1. We define the correlation coefficient between two nonzero vectors u, v in a inner product space $(V, \langle -, - \rangle, || - ||)$ as

$$\rho(u, v) = \frac{\langle u, v \rangle}{||u||||v||}$$

Example 5.2. If we use covariance to define an inner product for $RV((\Omega, \mathcal{F}, P), (\mathbb{R}, \mathcal{B}(\mathbb{R})))$ as

$$\langle X, Y \rangle = \operatorname{cov}(X, Y)$$

then

1. (symmetry) $\langle X, Y \rangle = \langle Y, X \rangle$.
2. (linearity) $\langle aX + bY, Z \rangle = a\langle X, Z \rangle + b\langle Y, Z \rangle$.
3. (positive semidefiniteness) $\langle X, X \rangle \geq 0$ with equality iff X is constant.

Therefore, covariance is not quite an inner product for $RV((\Omega, \mathcal{F}, P), (\mathbb{R}, \mathcal{B}(\mathbb{R})))$.

Example 5.3. It is an inner product for $RVZF((\Omega, \mathcal{F}, P), (\mathbb{R}, \mathcal{B}(\mathbb{R})))$. This inner product induces the norm

$$\begin{aligned} ||X|| &= \sqrt{\langle X, X \rangle} \\ &= \sqrt{\operatorname{cov}(X, X)} \\ &= \sqrt{\operatorname{var}(X)} \\ &= \sigma_X \end{aligned}$$

and the correlation coefficient

$$\rho(X, Y) = \operatorname{corr}(X, Y)$$

in probability theory.

Example 5.4. (averaging may reduce overfitting) Suppose we want to approximate a function

$$X \xrightarrow{f} Y$$

with a hypothesis. Suppose all have zero mean and finite variance.

- if we use one hypothesis h then we have
$$\operatorname{var}(h)$$
- if we use an average of hypotheses $\frac{h_1+\cdots+h_n}{n}$ then we have
$$\begin{aligned}
\operatorname{var}\left(\frac{h_1+\cdots+h_n}{n}\right) &= \operatorname{cov}\left(\frac{h_1+\cdots+h_n}{n}, \frac{h_1+\cdots+h_n}{n}\right) \\
&= \left\langle \frac{h_1+\cdots+h_n}{n}, \frac{h_1+\cdots+h_n}{n} \right\rangle \\
&= \frac{1}{n^2}\sum_{i,j=1}^{n} \langle h_i, h_j \rangle \\
&= \frac{1}{n^2}\sum_{i,j=1}^{n} \operatorname{cov}(h_i, h_j)
\end{aligned}$$
- if h_i, h_j are uncorrelated or independent then we have
$$\begin{aligned}
\operatorname{var}\left(\frac{h_1+\cdots+h_n}{n}\right) &= \frac{1}{n^2}\sum_{i=1}^{n} \operatorname{cov}(h_i, h_i) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \operatorname{var}(h_i)
\end{aligned}$$
- if h and $h_1, \ldots, h_n$ have similar variances then we have
$$\begin{aligned}
\operatorname{var}\left(\frac{h_1+\cdots+h_n}{n}\right) &\approx \frac{1}{n^2}\sum_{i=1}^{n} \operatorname{var}(h) \\
&= \frac{n\operatorname{var}(h)}{n^2} \\
&= \frac{\operatorname{var}(h)}{n}
\end{aligned}$$

6. Angle

A correlation coefficient induces angle.

Definition 6.1. We define the angle between two nonzero vectors u, v in an inner product space $(V, \langle -, - \rangle, || - ||, \rho(-, -))$ as

$$\angle(u, v) = \arccos(\rho(u, v))$$

Example 6.2. Via inner product and its induced norm and correlation coefficient, we get to define angle between vectors in $\mathbb{R}^m$ or random variables in $RVZF((\Omega, \mathcal{F}, P), (\mathbb{R}, \mathcal{B}(\mathbb{R})))$.