Lecture-02: Review of linear algebra

1 Linear algebra

1.1 Vector Space

Definition 1.1 (Vector addition). A set V is set to be equipped with vector addition mapping $+ : V \times V \to V$ defined by $+(v,w) = v + w$ for any two elements $v,w \in V$, if this mapping satisfies the following four axioms.

1. **Associativity:** For all vectors $u,v,w \in V$, we have $u + (v + w) = (u + v) + w$.
2. **Commutativity:** For all vectors $u,v \in V$, we have $u + v = v + u$.
3. **Additive identity:** There exists a zero vector $0 \in V$, such that $u + 0 = u$ for all $u \in V$.
4. **Additive inverse:** For each vector $u \in V$, there exists an additive inverse $-u \in V$ such that $u + (-u) = 0$.

Definition 1.2 (Scalar multiplication). A set V equipped with vector addition $+ : V \times V \to V$ is also equipped with field scalar multiplication mapping $\cdot : \mathbb{F} \times V \to V$ defined by $\cdot(\alpha,v) = \alpha v \in V$, if this mapping satisfies the following four axioms.

1. **Field compatibility:** For all scalars $\alpha,\beta \in \mathbb{F}$ and vector $u \in V$, we have $\alpha(\beta u) = (\alpha\beta)u$.
2. **Multiplicative identity:** There exists a multiplicative identity element $1 \in \mathbb{F}$, such that $1u = u$ for all $u \in V$.
3. **Distributivity over vector addition:** For each scalar $\alpha \in \mathbb{F}$ and vectors $u,v \in V$, we have $\alpha(v + u) = \alpha u + \alpha v$.
4. **Distributivity over field addition:** For all scalars $\alpha,\beta \in \mathbb{F}$ and vector $u \in V$, we have $(\alpha + \beta)u = \alpha u + \beta u$.

Definition 1.3. A vector space over the field $\mathbb{F}$ is a set V equipped with vector addition $+ : V \times V \to V$ and scalar multiplication $\cdot : \mathbb{F} \times V \to V$.

Definition 1.4. A set of vectors $W \subseteq V$ is called linearly independent, if for any nonzero vector $\alpha \in \mathbb{F}^W$ with finite $\sum_w \alpha_w$, we have $\sum_{w\in W} \alpha_w w \neq 0 \in V$.

Definition 1.5. The span of a set of vectors $W \subseteq V$ is defined by

$$\text{Span}(W) \triangleq \left\{ \sum_{w\in W} \alpha_w w : \alpha \in \mathbb{R}^W, \sum_{w\in W} \alpha_w \text{ finite} \right\}.$$

Definition 1.6. A basis of any vector space V, is a spanning set of linearly independent vectors, i.e., a linearly independent subset $W \subseteq V$ is basis of vector space V, if $\text{Span}(W) = V$.

Theorem 1.7. *All bases of a vector space V have identical cardinality, and defined to be its dimension.*

Example 1.8 (Vector space). Following are some common examples of vector spaces.

(a) Euclidean space of d-dimensions, denoted by $\mathbb{R}^d$.
(b) Space of continuous functions over a compact subset $[a,b]$ denoted by $C([a,b])$.
(c) Space of random variables defined over probability space $(\Omega,\mathcal{F},P)$ with finite pth moment denoted by L^p.

1.2 Inner Product Space

Definition 1.9. A *inner product space* is a vector space equipped with an inner product denoted by $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{R}$ that satisfies the following axioms.

1. **Symmetry:** For all vectors $u,v \in V$, we have $\langle u,v \rangle = \langle v,u \rangle$.
2. **Linearity:** For all scalars $\alpha,\beta \in \mathbb{F}$ and vectors $u,v,w \in V$, we have $\langle \alpha u + \beta v, w \rangle = \alpha \langle u,w \rangle + \beta \langle v,w \rangle$.
3. **Definiteness:** For all vectors $u \in V$, we have $\langle u,u \rangle \geqslant 0$, and $\langle u,u \rangle = 0$ iff $u = 0$.

Example 1.10 (inner product spaces). Following vector spaces are some common examples of inner product spaces.

(a) For the vector space $V = \mathbb{R}^d$ of d-dimensional vectors, the inner product is defined as $\langle u,v \rangle \triangleq u^T v = \sum_{i=1}^d u_i v_i$.

(b) For vector space $V = C(\mathbb{R}^d)$ of continuous functions, the inner product is defined as $\langle f,g \rangle \triangleq \int_{\mathbb{R}^d} (f,g)(t) dt$.

(c) For the vector space of random variables, the inner product $\langle \cdot,\cdot \rangle : L^p \times L^q \to \mathbb{R}$ is defined as $\langle X,Y \rangle \triangleq \mathbb{E}XY$ for conjugate pairs $p,q \geqslant 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$.

1.3 Norms

Definition 1.11. Norm is a mapping $\|\cdot\| : V \to \mathbb{R}_+$ that satisfy the following axioms.

1. **Definiteness:** For all vectors $v \in V$, we have $\|v\| = 0$ iff $v = 0$.
2. **Homogeneity:** For all scalars $\alpha \in \mathbb{R}$ and vectors $v \in V$, we have $\|\alpha v\| = |\alpha| \|v\|$.
3. **Triangle inequality:** For all vectors $v,w \in V$, we have $\|v + w\| \leqslant \|v\| + \|w\|$.

Example 1.12 (Norms). Let $p \geqslant 1$, then following are common examples of p-norms.

1. For a vector space $V = \mathbb{R}^d$, we can define the p-norm as $\|u\|_p \triangleq \left(\sum_{i=1}^d |u_i|^p \right)^{\frac{1}{p}}$ for all $u \in \mathbb{R}^d$.
2. For vector space $V = C(\mathbb{R}^d)$, the p-norm is defined as $\|f\|_p \triangleq \left(\int_{\mathbb{R}^d} |f|^p (t) dt \right)^{\frac{1}{p}}$ for all $f \in C(\mathbb{R}^d)$.
3. For vector space of random variables, the p-norm is defined as $\|X\|_p \triangleq \left(\mathbb{E} |X|^p \right)^{\frac{1}{p}}$ for all $X \in L^p$.

Example 1.13 (Special norms). Let $p \in \{1,2,\infty\}$, then following are common examples of p-norms.

1. Consider the vector space $V = \mathbb{R}^d$ and $x \in V$. For $p = 1$, we have $\|x\|_1 = \sum_{i=1}^d |x_i|$. For $p = 2$, the norm is Euclidean norm such that $\|x\|_2^2 = \langle x,x \rangle$. For $p = \infty$, we have $\|x\|_\infty = \max_i |x_i|$.
2. Consider the vector space $V = C(\mathbb{R}^d)$ and $f \in V$. For $p = 1$, we have $\|f\|_1 = \int_{t \in \mathbb{R}^d} |f| (t) dt$. For $p = 2$, the norm is Euclidean norm such that $\|f\|_2^2 = \langle f,f \rangle = \int_{t \in \mathbb{R}^d} |f|^2 (t) dt$. For $p = \infty$, we have $\|f\|_\infty = \sup_t |f| (t)$.
3. Consider the vector space V of random variables and $X \in V$. For $p = 1$, we have $\|X\|_1 = \mathbb{E}|X|$. For $p = 2$, the norm is Euclidean norm such that $\|X\|_2^2 = \langle X,X \rangle = \mathbb{E}X^2$. For $p = \infty$, we have $\|X\|_\infty = \sup_\omega |X| (\omega)$.

Proposition 1.14 (Holder's Inequality). *Let $p,q \geqslant 1$ be a conjugate pair, i.e. $\frac{1}{p} + \frac{1}{q} = 1$. Then,*

$$|\langle u,v \rangle| \leqslant \|u\|_p \|y\|_q \text{ for all } u,v \in \mathbb{R}^d.$$

Proof. The Holder's inequality is trivially true if $u = 0$ or $v = 0$. Hence, we assume that $\|u\|_p \|v\|_q > 0$, and let $a \triangleq \frac{|u_i|}{\|u\|_p}$ and $b \triangleq \frac{|v|_i}{\|v\|_q}$. We will use the Young's inequality $\frac{1}{p}a^p + \frac{1}{q}b^q \geqslant ab$ for all $a,b > 0$, that implies that

$$\frac{|u_i|^p}{p \|u\|_p^p} + \frac{|v_i|^q}{q \|v\|_q^q} \geqslant \frac{|u_i| |v_i|}{\|u\|_p \|v\|_q}, \text{ for all } i \in [d].$$

Since $|\langle u,v \rangle| \leqslant \sum_{i=1}^d |u_i| |v_i|$, we get the result by summing both sides over $i \in [d]$ in the above inequality.

Definition 1.15. For points $x, y \in \mathbb{R}^d$, the distance $d(x,y) \triangleq \|x-y\|^2$. Distance of a point x from a set is defined as $d(x,A) \triangleq \min\{d(x,y) : y \in A\}$.

Definition 1.16. For a vector $w \in \mathbb{R}^d$ and scalar $b \in \mathbb{R}$, we define a hyperplane as a set of points

$$E_{w,b} \triangleq \left\{ x \in \mathbb{R}^d : \frac{\langle w, x\rangle}{\|w\|} = -\frac{b}{\|w\|} \right\}.$$

1.4 Distance from hyperplanes

Lemma 1.17. *For a vector $w \in \mathbb{R}^d$ and $b \in \mathbb{R}$, we have $d(0, E_{w,b}) = |b| / \|w\|$.*

Proof. For a vector w, we define a unit vector $u \triangleq w/\|w\|$ such that $\|u\| = 1$ and $\langle w, u\rangle = \|w\|$. We define a vector $x_0 \triangleq -ub/\|w\|$ such that $\langle w, x_0\rangle = -b$. It follows that vector x_0 lies on the hyperplane $E_{w,b}$, is parallel to the unit vector u and at distance $d(0,x_0) = \|x_0\| = |b|/\|w\|$ from the origin. Consider a point $x \in E_{w,b}$, then from the linearity of inner products and definition of x_0, we obtain

$$\langle x - x_0, u\rangle = \langle x, u\rangle - \langle x_0, u\rangle = 0.$$

Hence, any point $x \in E_{w,b}$ on the hyperplane can be written as a sum of two orthogonal vectors $x = x_0 + x - x_0$. Therefore, $d(0,x)^2 = d(0,x_0)^2 + d(x_0,x)^2 \geqslant d(0,x_0^2)$, and hence $d(0, E_{w,b}) = d(0,x_0)$. ☐

Remark 1. A hyperplane $E_{w,b} = \left\{ x \in \mathbb{R}^d : \langle w, x\rangle + b = 0 \right\}$ is defined in terms of the unit vector $w/\|w\|$ and its distance $|b|/\|w\|$ from the origin.

Lemma 1.18. *The distance of any point $x \in \mathbb{R}^d$ to a hyperplane $E_{w,b}$ is given by $d(x, E_{w,b}) = \frac{|\langle w,x\rangle + b|}{\|w\|}$.*

Proof. Let $u = w/\|w\|$ be the unit vector in the direction of w and define $x_0 \triangleq -ub/\|w\|$. Any point y on a hyperplane $E_{w,b}$, can be written as sum of two orthogonal vectors $y = x_0 + y - x_0$. Any point $x \in \mathbb{R}^d$ can be represented as $x = \langle x, u\rangle u + v$, such that $\langle v, w\rangle = 0$. Therefore,

$$d(x, E_{w,b})^2 = \min_{y \in E_{w,b}} d(x,y)^2 = \min_{y \in E_{w,b}} d(x_0 + y - x_0, \langle x, u\rangle u + v)^2 \geqslant \left(\frac{\langle x, w\rangle + b}{\|w\|}\right)^2.$$

☐

Remark 2. The distance of a point $x \in \mathbb{R}^d$ from the hyperplane $E_{w,b}$ is given by $d(x, E_{w,b})$. If $\langle w, x\rangle + b > 0$, then the point x lies above the hyperplane $E_{w,b}$, and if $\langle w, x\rangle + b < 0$, then point x lies below the hyperplane $E_{w,b}$.