Lecture Notes on Machine Learning: Convex Sets

Technical Report · January 2019

Christian Bauckhage
University of Bonn

Lecture Notes on Machine Learning
Convex Sets

Christian Bauckhage
B-IT, University of Bonn

The sole and prosaic purpose of this crib sheet is to list (and illustrate) important terms and concepts related to convex sets.

Setting the Stage

Many algorithms for data mining, pattern recognition, and machine learning exploit properties of convex sets and convex functions. Data scientists should therefore be familiar with the basics of convexity.

In this note, we focus on convex sets and summarize definitions and results in this context. Since the concept of convexity appears all over mathematics[1], we will not cover all its manifestations. Instead, we confine our discussion to sets in the Euclidean space $\mathbb{R}^m$.

[1] R.J. Dwilewicz. A short history of convexity. *Differential Geometry – Dynamical Systems*, 11:112–129, 2009

To simplify our discussion, we first recall the notion of a convex combination of vectors and introduce some linear algebraic notation.

Consider a finite set of vectors $\mathcal{X} = \{x_1, x_2, \ldots, x_n\} \subset \mathbb{R}^m$. A vector $\boldsymbol{x} \in \mathbb{R}^m$ is called a ***convex combination*** of the vectors in $\mathcal{X}$, if it is a linear combination

$$\boldsymbol{x} = \sum_{j=1}^{n} w_j \, \boldsymbol{x}_j \tag{1}$$

where the coefficients $w_j \in \mathbb{R}$ obey the following constraints

$$w_j \geq 0 \tag{2}$$

$$\sum_{j=1}^{n} w_j = 1. \tag{3}$$

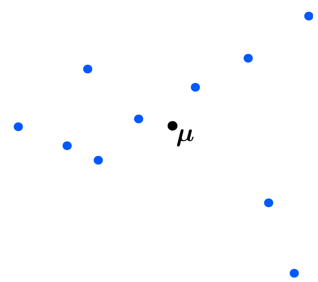

Figure 1: Given a set $\{x_1, x_2, \ldots, x_n\}$ of data points, the sample mean

$$\boldsymbol{\mu} = \frac{1}{n} \sum_{j=1}^{n} \boldsymbol{x}_j$$

is an example of a convex combination where each coefficient $w_j = \frac{1}{n}$.

Looking at (2) and (3), we realize that, in order for a set $\{w_1, \ldots, w_n\}$ of coefficients to be able to meet these constraints at all, each w_j must necessarily reside in the interval $[0, 1]$, that is $0 \leq w_j \leq 1$ for all j.

In order to condense the expression in (1), we may collect the vectors $\boldsymbol{x}_j$ in a matrix $\boldsymbol{X} = [x_1, x_2, \ldots, x_n] \in \mathbb{R}^{m \times n}$ and the coefficients w_j in a vector $\boldsymbol{w} \in \mathbb{R}^n$ and write

$$\boldsymbol{x} = \boldsymbol{X}\boldsymbol{w}. \tag{4}$$

To express that $\boldsymbol{w}$ is a non-negative vector whose components obey the inequalities in (2), we henceforth write

$$\boldsymbol{w} \succeq \boldsymbol{0} \tag{5}$$

where $\boldsymbol{0}$ is the vector of all zeros. To abbreviate the sum-to-one constraint in (3), we henceforth write it as an inner product

$$\boldsymbol{1}^\intercal \boldsymbol{w} = 1 \tag{6}$$

where $\boldsymbol{1}$ is the vector of all ones. Finally, any vector $\boldsymbol{w}$ that has the two properties in (5) and (6) is called a ***stochastic vector***.

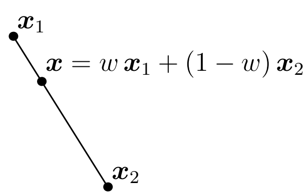

Figure 2: Given two points x_1 and x_2, the set of convex combinations

$$\left\{ x = w\,x_1 + (1-w)\,x_2 \;\middle|\; w \in [0,1] \right\}$$

defines the line segment $\overline{x_1\,x_2}$.

Convex Sets

A set $\mathcal{S} \subseteq \mathbb{R}^m$ is called a ***convex set***, if every point on the line segment between any two points in $\mathcal{S}$ is also in $\mathcal{S}$, that is if

$$\forall \, \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{S} \ \wedge \ \forall \, w \in [0,1] : w \, \boldsymbol{x}_1 + (1-w) \, \boldsymbol{x}_2 \in \mathcal{S}.$$

The empty set $\emptyset$ and the whole space $\mathbb{R}^m$ are convex; subspaces and (open) half-spaces of $\mathbb{R}^m$ are convex; if $\{\mathcal{S}_i\}_{i=1}^n$ is a set of convex sets, their intersection

$$\mathcal{S} = \bigcap_{i=1}^{n} \mathcal{S}_i \tag{7}$$

is convex.

An ***extreme point*** of a convex set $\mathcal{S}$ is any point $\boldsymbol{x} \in \mathcal{S}$ that is *not* a convex combination of other points in $\mathcal{S}$. In other words, if $\boldsymbol{x}$ is an extreme point, then $\boldsymbol{x} = w \, \boldsymbol{y} + (1-w) \, \boldsymbol{z}$ with $\boldsymbol{y}, \boldsymbol{z} \in \mathcal{S}$ and $w \in [0,1]$ implies that $\boldsymbol{x} = \boldsymbol{y} = \boldsymbol{z}$. Every compact convex set has at least one extreme point.

The ***convex hull*** $\mathcal{C}(\mathcal{S})$ of a set $\mathcal{S}$ is the set of all possible convex combinations of points in $\mathcal{S}$, that is

$$\mathcal{C}(\mathcal{S}) = \left\{ \sum_{\boldsymbol{x}_j \in \mathcal{R}} w_j \, \boldsymbol{x}_j \, \middle| \, \mathcal{R} \subseteq \mathcal{S}, \ |\mathcal{R}| < \infty, \ \boldsymbol{w} \succeq \boldsymbol{0}, \ \boldsymbol{1}^\intercal \boldsymbol{w} = 1 \right\}.$$

The ***Krein-Milman*** theorem establishes that every compact convex set is the convex hull of its extreme points.

A ***V-polytope*** is the convex hull of a set $\mathcal{X} = \{\boldsymbol{x}_j\}_{j=1}^n$ of finitely many points, namely

$$\mathcal{P} = \left\{ \sum_{j=1}^{n} w_j \, \boldsymbol{x}_j \, \middle| \, \boldsymbol{w} \succeq \boldsymbol{0}, \ \boldsymbol{1}^\intercal \boldsymbol{w} = 1 \right\}. \tag{8}$$

The finitely many extreme points of a V-polytope are called ***vertices***. The number k of vertices of $\mathcal{P}$ is bounded by the number n of points in $\mathcal{X}$, that is $k \leq n$. Also, by virtue of the Krein-Milman theorem, if

$$\mathcal{V} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\} \subset \mathcal{X} \subset \mathcal{P} \tag{9}$$

is the set of all vertices of $\mathcal{P}$, then every point $\boldsymbol{x} \in \mathcal{P}$ can be expressed as a convex combination of the points in $\mathcal{V}$. In other words, if we gather the vertices in a matrix $\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k] \in \mathbb{R}^{m \times k}$, then $\boldsymbol{x} = \boldsymbol{V}\boldsymbol{u}$ where $\boldsymbol{u} \in \mathbb{R}^k$ is a stochastic vector.

An ***H-polytope*** is a convex set that results from intersecting finitely many, say M, half-spaces and we have

$$\mathcal{P} = \left\{ \boldsymbol{x} \in \mathbb{R}^m \, \middle| \, \boldsymbol{a}_i^\intercal \boldsymbol{x} \leq b_i, \ 1 \leq i \leq M \right\}. \tag{10}$$

In other words, an H-polytope $\mathcal{P}$ is the set of all $\boldsymbol{x} \in \mathbb{R}^m$ such that

$$\boldsymbol{A}\boldsymbol{x} \preceq \boldsymbol{b} \tag{11}$$

where $\boldsymbol{A} \in \mathbb{R}^{M \times m}$ and $\boldsymbol{b} \in \mathbb{R}^M$. If all solutions are bounded in the sense that $\|\boldsymbol{x}\| < \infty$, then $\mathcal{P}$ is bounded.

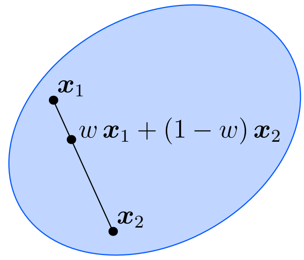

Figure 3: A convex set.

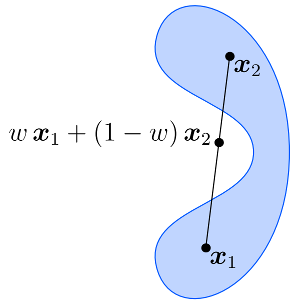

Figure 4: A non-convex set.

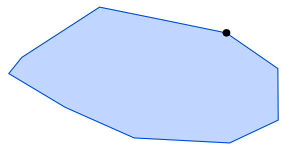

Figure 5: An extreme point or vertex.

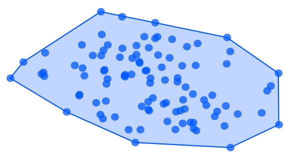

Figure 6: A V-polytope is the convex hull of a set of finitely many points.

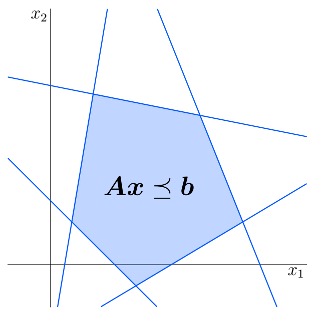

Figure 7: A bounded H-polytope.

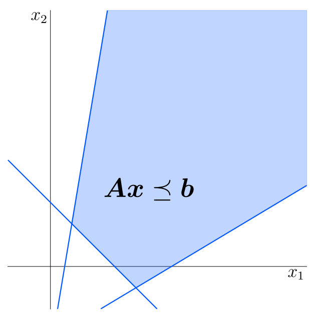

Figure 8: An unbounded H-polytope.

Each inequality in (10) defines a ***supporting hyperplane*** of $\mathcal{P}$. In general, if $\mathcal{S} \subseteq \mathbb{R}^m$ is a convex set and x_0 is a point on the boundary $\partial\mathcal{S}$ of $\mathcal{S}$, a supporting hyperplane of $\mathcal{S}$ through x_0 is a set

$$\mathcal{H} = \left\{ \boldsymbol{x} \mid \boldsymbol{a}^\intercal \boldsymbol{x} = \boldsymbol{a}^\intercal \boldsymbol{x}_0 \right\} \tag{12}$$

where $\boldsymbol{a} \neq \mathbf{0} \in \mathbb{R}^m$ and

$$\boldsymbol{a}^\intercal \boldsymbol{x} \leq \boldsymbol{a}^\intercal \boldsymbol{x}_0 \tag{13}$$

for all $\boldsymbol{x} \in \mathcal{S}$. The ***supporting hyperplane theorem*** states that, if $\mathcal{S} \subseteq \mathbb{R}^m$ is convex, there exists a supporting hyperplane for every boundary point $\boldsymbol{x}_0 \in \partial\mathcal{S}$. Vice versa, if $\mathcal{S}$ is a closed set with nonempty interior such that every point $\boldsymbol{x}_0 \in \partial\mathcal{S}$ has a supporting hyperplane, then $\mathcal{S}$ is convex.

The ***separating hyperplane theorem*** states that, if $\mathcal{S}_1, \mathcal{S}_2 \subset \mathbb{R}^m$ are two nonempty and disjoint convex sets, there exist $\boldsymbol{a} \neq \mathbf{0} \in \mathbb{R}^m$ and $b \in \mathbb{R}$ such that

$$\forall\, \boldsymbol{x} \in \mathcal{S}_1 : \boldsymbol{a}^\intercal \boldsymbol{x} \leq b \tag{14}$$

$$\forall\, \boldsymbol{x} \in \mathcal{S}_2 : \boldsymbol{a}^\intercal \boldsymbol{x} \geq b \tag{15}$$

and the plane $\mathcal{H} = \{\boldsymbol{x} \mid \boldsymbol{a}^\intercal \boldsymbol{x} = b\}$ is called the ***separating hyperplane***.

If there exists a separating hyperplane between two sets $\mathcal{S}_1$ and $\mathcal{S}_2$, they are said to be ***linearly separable***.

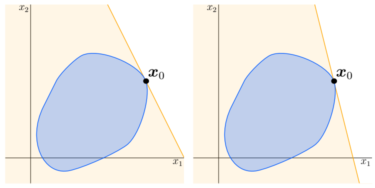

Figure 9: There may be more than one supporting hyperplane through a boundary point $\boldsymbol{x}_0 \in \partial\mathcal{S}$.

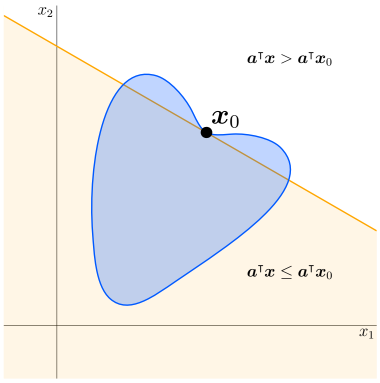

Figure 10: If $\mathcal{S}$ is not convex, there will be boundary a point $\boldsymbol{x}_0 \in \partial\mathcal{S}$ such that (13) does not hold for all $\boldsymbol{x} \in \mathcal{S}$.

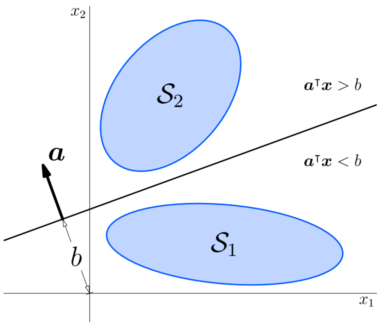

Figure 11: Linearly separable sets can be separated by at least one hyperplane $\mathcal{H} = \{\boldsymbol{x} \mid \boldsymbol{a}^\intercal \boldsymbol{x} = b\}$.

Given the notions of V- and bounded H-polytopes, one can prove the ***main theorem of polytope theory*** which states that V-polytopes and bounded H-polytopes are equivalent. Every V-polytope can be characterized in terms of a finite system of inequalities and every bounded H-polytope can be characterized in terms of the convex hull of a finite set of vertices.

This theorem has several notable consequences. For example, any intersection of a polytope with an affine subspace is yet another polytope. In fact, every image (or projection) $P(\mathcal{P})$ of a polytope $\mathcal{P}$ under an affine map

$$P(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} \tag{16}$$

is a polytope; in particular, every vertex of $P(\mathcal{P})$ is the image of a vertex of $\mathcal{P}$.

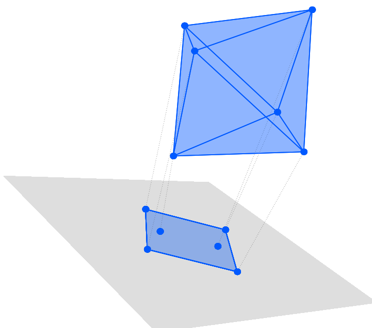

Figure 12: Any affine image or projection $P(\mathcal{P})$ of a polytope $\mathcal{P}$ is a polytope.

One can also show that every convex polytope is a projection of a higher dimensional simplex. The ***standard simplex*** Δ^{m-1} is the convex hull of the standard basis $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_m\} \subset \mathbb{R}^m$, that is

$$\Delta^{m-1} = \left\{ \sum_{i=1}^{m} w_i\, \boldsymbol{e}_i \,\middle|\, \boldsymbol{w} \in \mathbb{R}^m,\ \boldsymbol{w} \succeq \mathbf{0},\ \mathbf{1}^\intercal \boldsymbol{w} = 1 \right\}.$$

The standard simplex is thus a polytope whose vertices coincide with the standard basis vectors. Equivalently, the standard simplex Δ^{m-1} is the set of all stochastic vectors $\boldsymbol{w} \in \mathbb{R}^m$.

At first sight it seems strange that the standard simplex in $\mathbb{R}^m$ is called Δ^{m-1}. But consider this: although it lives in an m-dimensional space, a standard simplex is only an $m-1$ dimensional object.

For the case where $m = 3$, we can see this in Fig. 13. For the general case, we note that the elements w_j of $\boldsymbol{w} \in \Delta^{m-1}$ are not fully independent: because of the sum-to-one constraint $\mathbf{1}^\intercal \boldsymbol{w} = 1$, any element w_k of $\boldsymbol{w}$ can be expressed in terms of the remaining $m-1$ elements, namely

$$w_k = 1 - \sum_{k \neq i} w_i. \tag{17}$$

In other words, we only need $m-1$ variables to characterize a point in Δ^{m-1} and thus the ***intrinsic dimension*** of Δ^{m-1} is $m-1$.

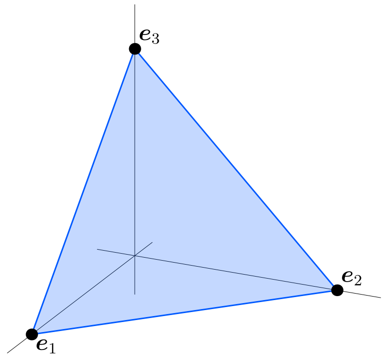

Figure 13: Standard simplex Δ^2 in $\mathbb{R}^3$.

Notes and Further Reading

Readers interested in formal proofs of our many claims in this crib-sheet might find them in the excellent books by Ziegler[2], Boyd and Vandenberghe[3], and Dattorro[4], the latter being freely available on the Web.

[2] G.M. Ziegler. *Lectures on Polytopes*. Springer, 1995

[3] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004

[4] J. Dattorro. *Convex Optimization & Euclidean Geometry*. Meboo Publishing, 2005

Acknowledgments

This material was prepared within project P3ML which is funded by the Ministry of Education and Research of Germany (BMBF) under grant number 01/S17064. The authors gratefully acknowledge this support.

References

S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

J. Dattorro. *Convex Optimization & Euclidean Geometry*. Meboo Publishing, 2005.

R.J. Dwilewicz. A short history of convexity. *Differential Geometry – Dynamical Systems*, 11:112–129, 2009.

G.M. Ziegler. *Lectures on Polytopes*. Springer, 1995.