ETH zürich

Math Recap

Jonas Hübotter
20.02.2026

Matrix Algebra and Geometric Constructs

$$Ax = \sum_i x_i a_i$$

$$\begin{bmatrix} | & & | \\ a_1 & \cdots & a_k \\ | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix}$$

algebraic

$$Bx = \begin{bmatrix} b_1^T x \\ \vdots \\ b_N^T x \end{bmatrix}$$

$$\begin{bmatrix} - & b_1^T & - \\ & \vdots & \\ - & b_N^T & - \end{bmatrix}$$

Geometric

Inner and Outer Products

Euclidean inner product: $\langle u, v \rangle = u^T v = \sum_i u_i \cdot v_i$

induces a norm: $\|u\|_2 = \sqrt{\langle u, u \rangle} = \sqrt{\sum_i u_i^2}$

Outer product: $uv^T = \begin{bmatrix} | & & | \\ u v_1 & \cdots & u v_n \\ | & & | \end{bmatrix} \in \mathbb{R}^{m \times n}$ for $u \in \mathbb{R}^m, v \in \mathbb{R}^n$

Singular Value Decomposition

Theorem: Let $A \in \mathbb{R}^{m \times n}$. We can decompose A as a product $A = U \Sigma V^T$, where $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthogonal, and Σ is a $m \times n$ diagonal matrix with non-negative diagonal entries.

$$A = U \Sigma V^T$$

$$= \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \underset{\uparrow}{\sigma_m} \end{bmatrix} \begin{bmatrix} - & v_1^T & - \\ & \vdots & \\ - & v_n^T & - \end{bmatrix}$$

$\sigma_i \geq 0$

$$\langle u_i, u_j \rangle = \begin{cases} 1 & , \text{if } i = j \\ 0 & , \text{otherwise.} \end{cases}$$

$$\langle v_i, v_j \rangle = \begin{cases} 1 & , \text{if } i = j \\ 0 & , \text{otherwise} \end{cases}$$

Singular Value Decomposition: Geometric Interpretation

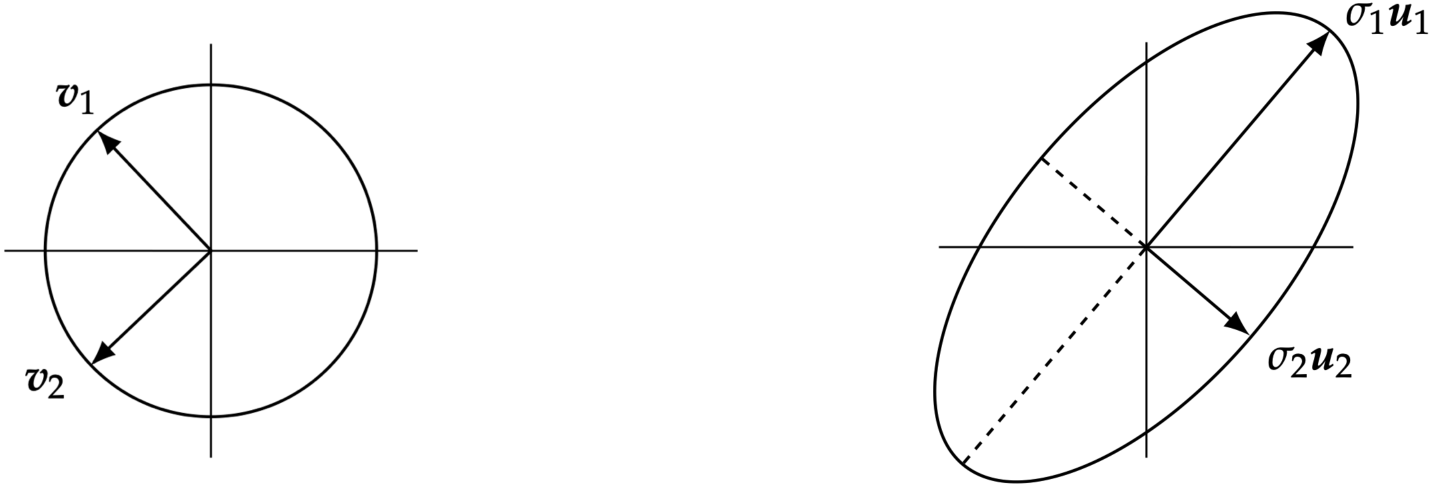
$$Ax = U\Sigma V^T x$$

reflection/rotation

stretch/squeeze

rotation/reflection

Singular Value Decomposition: Illustration



Orthogonal Projections

Consider linearly independent vectors $\underbrace{a_1, \ldots, a_n}$ $\in \mathbb{R}^m$ $(n \leq m)$

$A = \begin{bmatrix} | & & | \\ a_1 & \cdots & a_n \\ | & & | \end{bmatrix} \in \mathbb{R}^{m \times n}$

$\text{rank}(A) = n$ (full-rank)

Q: What is the orthogonal projection y of a vector x onto the $\text{span}(\{a_1, \ldots, a_n\})$?

Requirements

1) $y \in \text{span}(\{a_1, \ldots, a_n\}) \Rightarrow y = Av$

2) $x - y \perp \text{span}(\{a_1, \ldots, a_n\}) \Leftrightarrow \langle x-y, w \in \text{span}(\{a_1, \ldots, a_n\}) \rangle = 0$

$\Leftrightarrow \langle x-y, a_i \rangle = 0$

$\Leftrightarrow A^T(x-y) = 0$

Orthogonal Projections: Continued

From 2) $A^T(x-y)=0 \iff A^Tx = A^Ty$

$= \underbrace{A^TA}_{\text{invertible since } a_1,\dots,a_n \text{ are linearly independent}} v$ (from 1)

$\Rightarrow v = (A^TA)^{-1}A^T x$

From 1) $y = Av = \underbrace{A(A^TA)^{-1}A^T}_{\text{projection matrix}} x =: \mathrm{Proj}_A(x)$

Eigenvalues, Eigenvectors

Let $\lambda \in \mathbb{R}$, $v \in \mathbb{R}^n$, $v \neq 0$. Let $A \in \mathbb{R}^{n \times n}$

$$A v = \underbrace{\lambda}_{\text{eigenvalue}} \overbrace{v}^{\text{eigenvector}}$$

Finding them: find roots of characteristic equation $\det(A - \lambda I) = 0$

If A is symmetric ($A^T = A$), then all eigenvalues are real.

Relationship to SVD: $\sigma_i^2(A) = \lambda_i(A^T A) = \lambda_i(A A^T)$

Spectral decomposition: $A = X \Lambda X^{-1}$

$$X = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix} \qquad \Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

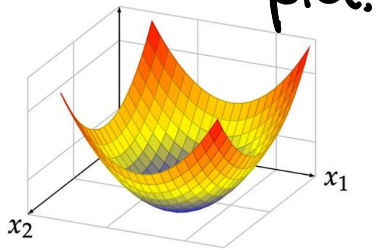
Quadratic Forms

$x \mapsto x^T A x$

Let $A \in \mathbb{R}^{n \times n}$ be symmetric

→ quadratic form: $x^T A x = \sum_i \sum_j a_{ij} \cdot x_i \cdot x_j$

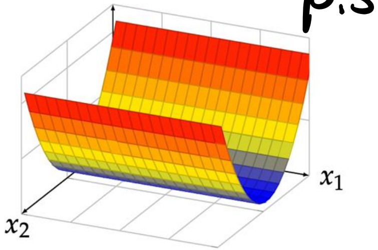
p.d.



A is positive definite (p.d.) if $x^T A x > 0 \quad \forall x \neq 0$

$\Leftrightarrow \lambda_i > 0 \quad \forall i = 1, \dots, n$

p.s.d



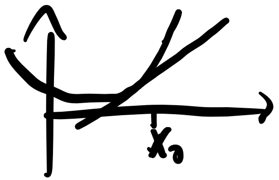
A is positive semi-definite (p.s.d.) if $x^T A x \geq 0 \quad \forall x \neq 0$

$\Leftrightarrow \lambda_i \geq 0 \quad \forall i = 1, \dots, n$

Cholesky decomposition (square root): If A p.s.d., then A can be factored as $A = LL^T$, where L is a lower-triangular matrix. $L = \begin{bmatrix} & 0 \\ & \end{bmatrix}$

Multivariate Derivatives

$f: \mathbb{R} \to \mathbb{R}$



$f: \mathbb{R}^n \to \mathbb{R}^m$

What is the best linear approximation?

$$f(x_0+\varepsilon) \approx f(x_0) + \underbrace{\underbrace{Df(x_0)}_{\text{linear map / matrix}}[\varepsilon]}_{\text{matrix-vector product}}$$

"Jacobian"

$$[Df(x_0)]_{ij} = \frac{\partial f_i}{\partial x_j}(x_0)$$

If $m=1$, then $(Df(x_0))^T = \nabla f(x_0)$ "gradient"

$Df(x_0) \in \mathbb{R}^{m \times n}$ $\quad (\mathbb{R}^n \to \mathbb{R}^m)$

$\nabla f(x_0) \in \mathbb{R}^n$

$m=1$:

derivative: $1 \times n$

gradient: $n \times 1$

Chain Rule

$f: \mathbb{R}^n \to \mathbb{R}^m, \quad g: \mathbb{R}^k \to \mathbb{R}^n$

linear map

$$D(f \circ g)(x_0) = Df(\underbrace{g(x_0)}_{\text{const}}) \circ Dg(x_0) = Df(g(x_0))\, Dg(x_0)$$

Example: vectors

$g(t) = x_0 + t \cdot v$

$$\frac{d}{dt} f(x_0 + t \cdot v)\Big|_{t=0} = D(f \circ g)(0) = \underbrace{Df(\underbrace{g(0)}_{x_0})}_{\in \mathbb{R}^{1\times n}} \underbrace{Dg(0)}_{v \,\in \mathbb{R}^n}$$

scalar

$$= \left[\frac{\partial f}{\partial x_1}(x_0) \; \dots \; \frac{\partial f}{\partial x_n}(x_0) \right] v$$

$g(0)$

$$= (\nabla f(x_0))^T v = \langle \nabla f(x_0), v \rangle$$

Taylor Expansion and Second-order Derivatives

$f: \mathbb{R}^n \to \mathbb{R}$

first-order expansion at x_0:

lecture notes

$$f(x) = f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + o(\|x - x_0\|)$$

second-order expansion at x_0:

$$f(x) = f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \frac{1}{2}(x - x_0)^T D^2 f(x_0)(x - x_0) + o(\|x - x_0\|^2)$$

$$\underbrace{D^2 f(x_0)}_{\text{matrix}} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x_0) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(x_0) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x_0) \end{bmatrix}$$

"Hessian"

Random Variables, Distributions and Densities

Probability space $(\Omega, \mathcal{F}, \mathbb{P})$

Ω: sample space (atomic events)

$\mathcal{F}$: event space $\mathcal{F} \subseteq 2^{\Omega}$

$\mathbb{P}$: probability function: $\mathbb{P}: \mathcal{F} \to [0,1]$

Kolmogorov axioms:

1) $0 \leq \mathbb{P}(A) \leq 1 \quad \forall A \in \mathcal{F}$
2) $\mathbb{P}(\Omega) = 1$
3) For a countable mutually disjoint set of events $\{A_i \in \mathcal{F}\}_i$ $(A_i \cap A_j = \emptyset)$
$\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$

Random variable: $X: \Omega \to \mathbb{R}$

deterministic for given input $\omega \in \Omega$

distribution: $\mathbb{P}_X(A) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\})$

subset of values $A \subseteq \mathbb{R}$

cumulative distribution function (CDF) $F_X(x) = \mathbb{P}((-\infty, x]) = \mathbb{P}(X \leq x)$

probability density function (PDF) $p_X: \mathbb{R} \to [0, \infty)$, such that for any $I \subseteq \mathbb{R}$

$$\mathbb{P}(I) = \int_I p_X(x)\,dx$$

Joint Distributions

Let X, Y be two random variables on the same $(\Omega, \mathcal{F}, \mathbb{P})$

joint distribution: $\mathbb{P}_{X,Y}(A) = \mathbb{P}(\{\omega \in \Omega : (X(\omega), Y(\omega)) \in A\})$, for $A \in \mathbb{R}^2$

joint CDF: $F_{X,Y}(x,y) = \mathbb{P}_{X,Y}((-\infty, x] \times (-\infty, y]) = \mathbb{P}(X \leq x, Y \leq y)$

joint PDF: $p_{X,Y}(x,y)$, such that $\mathbb{P}_{X,Y}(A) = \int_A p_{X,Y}(x,y)\, dx\, dy$

Marginal Distributions

Given joint distribution $\mathbb{P}_{X,Y}$, what is the distribution $\mathbb{P}_X$?

"marginal" distribution: $\mathbb{P}_X(A) = \mathbb{P}_{X,Y}(A \times \mathbb{R})$ for $A \subseteq \mathbb{R}$

given the joint PDF $p_{X,Y}$, marginal PDF p_X:

$$p_X(x) = \int_{\mathbb{R}} p_{X,Y}(x,y)\, dy$$

Conditional Distributions and Independence

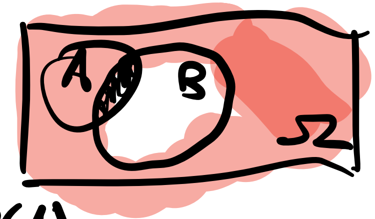
Let's consider two events A, B.

events A and B are independent iff $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$

$$p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y)$$

if $\mathbb{P}(B) > 0$, the conditional probability of A given B is

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$



Bayes' rule: $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B \mid A) \cdot \mathbb{P}(A)}{\mathbb{P}(B)}$

Expected Value

Expectation: $\mathbb{E}[X] = \int_{\mathbb{R}} x \cdot p_X(x)\, dx$

Variance: $\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

If X, Y independent, then $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$

Conditional expectation $\mathbb{E}[X \mid Y = y] = \int_{\mathbb{R}} x \cdot p_{X|Y}(x|y)\, dx$