Math for ML

Tanvir

November 2025

Matrix multiplication

Element-wise

$$A^{m \times n} \; B^{n \times p} = C^{m \times p}$$

$$c_{i,j} = (\cdots a_i \cdots) \cdot \begin{pmatrix} \cdot \\ \cdot \\ b_j \\ \cdot \\ \cdot \end{pmatrix}$$

```
np.einsum('ik,kj->ij', A, B)
```

Column-wise

$$\begin{pmatrix} \cdot \\ \cdot \\ c_k \\ \cdot \\ \cdot \end{pmatrix} = A \begin{pmatrix} \cdot \\ \cdot \\ b_k \\ \cdot \\ \cdot \end{pmatrix}$$

Row-wise

$$(\cdots c_k \cdots) = (\cdots a_k \cdots) \; B$$

Outer-product

$$C = \sum_{i=1}^{n} \begin{pmatrix} \cdot \\ \cdot \\ a_i \\ \cdot \\ \cdot \end{pmatrix} (\cdots b_i \cdots)$$

Block-wise

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$$

$$C = AB = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = \begin{pmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{pmatrix}.$$

Column picture: $Ax = b$

$$A\,x = b$$

$$x_1 \begin{pmatrix} \cdot \\ \cdot \\ a_1 \\ \cdot \\ \cdot \end{pmatrix} + x_2 \begin{pmatrix} \cdot \\ \cdot \\ a_2 \\ \cdot \\ \cdot \end{pmatrix} = b$$

Notes:

- Linear combination of columns

Row picture: $xA = B$

$$x \begin{pmatrix} \cdots r_1 \cdots \\ \cdots r_2 \cdots \\ \cdots r_3 \cdots \end{pmatrix} = \begin{pmatrix} x_1\, r_1 + x_2\, r_2 + x_3\, r_3 \end{pmatrix}$$

Notes:

- Linear combination of rows

Elimination: $A \to U$

$$E_n \cdots E_2 E_1 A = U$$

$$A\,x = b$$
$$E_n \cdots E_2 E_1 A\,x = E_n \cdots E_2 E_1 b$$
$$U\,x = E_n \cdots E_2 E_1 b$$

Practice: Solve the system of equations:

$$x - y - z + u = 0$$
$$2x + 2z = 8$$
$$-y - 2z = -8$$
$$3x - 3y - 2z + 4u = 7$$

Matrix inverse

Square matrix

$A^{-1} A = I = A A^{-1}$

$A x = b$
$\quad x = A^{-1} b$

$(AB)^{-1} = B^{-1} A^{-1}$

$\left(A^{\mathsf{T}}\right)^{-1} = \left(A^{-1}\right)^{\mathsf{T}}$

Notes:

- $\det(A) = 0$ implies that A is not invertible or A is singular.

- For $x \neq 0$, $A x = 0$ implies that A is singular.

- The Gauss-Jordan elimination finds the inverse: $E(A|I) = (I|A^{-1})$.

Practice: Conditions for inverse of $\begin{pmatrix} a & b & b \\ a & a & b \\ a & a & a \end{pmatrix}$ and the inverse when it exists.

Matrix Factorization

$A = L U$

$E A = U$
$\quad A = E^{-1} U$

$L = E^{-1}$

Notes:

- For $A^{n \times n}$, $\mathcal{O}(n^3)$.

- U has row echelon (staircase) form. In reduced row echelon form, $\mathrm{rref}(A)$, the pivots are 1's and above and below each pivot there are 0's.

Practice: When does LU factorization exist and what are the factors for $\begin{pmatrix} 1 & 0 & 1 \\ a & a & a \\ b & b & a \end{pmatrix}$.

Permutation matrices

Permutation matrices do row exchanges. There are $n!$ permutation matrices and they form a group.

$$P^{-1} = P^T$$

With row exchanges,

$$PA = LU$$
$$A = P^T LU$$

Symmetric matrix

$$A^T = A$$

Notes:

- $A^T A$ is square and symmetric.

Vector space

Set of vectors closed under: (1) vector addition (2) scalar multiplication. Example: $\mathbb{R}^2$. The set $\{u \in \mathbb{R}^2 : u_i > 0\}$ is not a vector space.

Vector subspace

A vector space within a larger vector space.

Notes:

- A line through the origin is a subspace in $\mathbb{R}^2$. These vectors have two components, so the line-subspace is not the same as $\mathbb{R}^1$.

- The set with the zero vector is a subspace in $\mathbb{R}^2$: $\left\{ (0,0)^T \right\}$.

- If S and T are subspaces, $S \cap T$ is a subspace. $S \cup T$ may not be a subspace.

Practice: $x_1 = \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix}$, $x_2 = \begin{pmatrix} 2 \\ 4 \\ 0 \end{pmatrix}$

1. Find V_1 = Subspace generated by x_1; V_2 = Subspace generated by x_2. Describe $V_1 \cap V_2$.

2. Find V_3 = Subspace generated by $\{x_1, x_2\}$. Is $V_3 = V_1 \cup V_2$? Find a subspace S of V_3 such that $x_1 \notin S$, $x_2 \notin S$.

3. What is $V_3 \cap \{x - y \text{ plane}\}$?

Practice: Which are subspaces of $\mathbb{R}^3 = \left\{ \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \right\}$?

1. $b_1 + b_2 - b_3 = 0$

2. $b_1 b_2 - b_3 = 0$

3. $\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + c_1 \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} + c_2 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$

4. $\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + c_1 \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} + c_2 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$

Matrix subspaces

Columnspace

$C(A)$: Linear combinations of columns of A.
$Ax = b$ has a solution when $b \in C(A)$.

$$C\left(\begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{bmatrix}\right) = C\left(\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix}\right)$$

The column space of $\begin{pmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{pmatrix}$ is a two-dimensional subspace in $\mathbb{R}^4$.

Nullspace

$N(A)$: All solutions of $Ax = 0$.
We can use elimination to compute $N(A)$. $EA = U$. The number of pivots in U is $\text{rank}(A)$.

With $A = \begin{pmatrix} 1 & 2 & 2 & 2 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 8 & 10 \end{pmatrix}$ and $U = \begin{pmatrix} 1 & 2 & 2 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix}$, we now solve $Ux = 0$.

Here x_1, x_3 are pivot variables and x_2 and x_4 are free variables. We can have one special solution for each free variable by setting that free variable to 1 and the rest of the free variables to 0. $N(A)$ would then be linear combinations of the special solutions. For $A^{m \times n}$, there are $n - \text{rank}(A)$ free variables.

$$\text{rref}(A) = \begin{pmatrix} 1 & 2 & 0 & -2 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix} \text{ has the form } \begin{pmatrix} I & F \\ 0 & 0 \end{pmatrix} \text{ with I as the pivots and F as}$$

the free's. Special solutions are the columns of $\begin{pmatrix} -F \\ I \end{pmatrix}$.

$$N\left(\begin{bmatrix} 1\,1\,2 \\ 2\,1\,3 \\ 3\,1\,4 \\ 4\,1\,5 \end{bmatrix} \right) \text{ is a subspace of } \mathbb{R}^3.$$

If $b \neq 0$, the set of solutions of $Ax = b$ is not a subspace.
$\text{rank}(A) = \text{rank}(A^T)$.
If A's columns are independent, $N(A)$ is empty.

Rowspace: $C(A^T)$

The space spanned by the row vectors. Same as $C(A^T)$. $\text{rank}(C(A)) = \text{rank}(C(A^T))$, because each row can have at most 1 pivot, each column can have at most 1 pivot. Also, a row is a constraint. Each new constraint reduces the number of free variables by 1 and thus increases pivot count by 1.

Left nullspace: $N(A^T)$

For $A^{m \times n}$, $\text{rank}(N(A^T)) = m - \text{rank}(C(A))$.

Practice: Suppose

$$B = \begin{pmatrix} 1 & & \\ 2 & 1 & \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 5 & 0 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

Find a basis for and compute the dimension of each of the 4 fundamental subspaces.
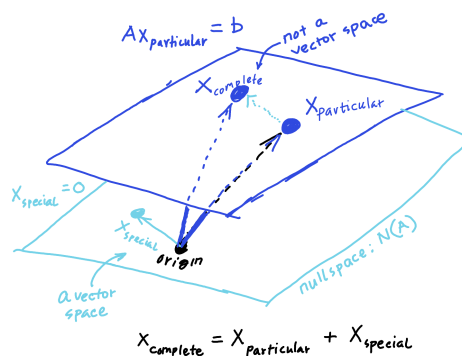
Solve $Ax = b$

Solvable only if $b \in \text{col}(A)$. To solve, use Gaussian elimination.
$x_{\text{complete}} = x_{\text{particular}} + x_{\text{special}}$.
    Say $Ax_1 = b$ and $Ax_2 = b$. Then, $A(x_2 - x_1) = 0$. So, difference between any two solutions, Δx, is in nullspace: $A \Delta x = 0$. Note, $x_2 = x_1 + \Delta x$.
    We get $x_{\text{particular}}$ by setting all free variables to 0. We get one special nullspace solution per free variable. Linear combinations of x_{special}'s is the nullspace.

$$X_{complete} = X_{particular} + X_{special}$$

Practice: The set S of points $P(x, y, z)$ such that $x - 5y + 2z = 9$ is a _____ in $\mathbb{R}^3$. It is _____ to the _____ S_0 of $P(x, y, z)$ such that $x - 5y + 2z = 0$. All points of S have the form:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + c_1 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Rank

For $A^{m \times n}$, the rank r is the number of pivots during elimination. $r \leqslant m$, $r \leqslant n$.

Full column rank, $r = n < m$

0 or 1 solution.

No free variables. $N(A) = \{0\}$. Solution to $Ax = b$ exists only if $b \in C(A)$ and in that case it is unique.
$\text{rref}(A) = \begin{pmatrix} I \\ 0 \end{pmatrix}$

Full row rank, $r = m < n$

∞ solutions.

There are $(n - m)$ free variables. $Ax = b$ has $x_{particular}$ only if $b \in C(A)$. There are $(n - m)$ $x_{special}$'s whose linear combinations give $N(A)$.
$\text{rref}(A) = \begin{pmatrix} I & F \end{pmatrix}$

Full rank, $r = m = n$

1 solution

$r = m = n$. There is always a unique solution to $Ax = b$. $N(A) = \{0\}$.
$\text{rref}(A) = I$.

Not-full rank, $r < m, r < n$

0 or ∞ solutions.
$$\text{rref}(A) = \begin{pmatrix} I & F \\ 0 & 0 \end{pmatrix}$$

Practice: Find all solutions, depending on b_1, b_2, b_3:

$$\begin{aligned} x - 2y - 2z &= b_1 \\ 2x - 5y - 4z &= b_2 \\ 4x - 9y - 8z &= b_3 \end{aligned}$$

Linearly independent vectors

The columns are linearly independent vectors if $N(A) = \{0\}$.
   Basis for a space is a set of linearly independent vectors that span the space. Given a space, every basis for the space has the same number of vectors and the number of vectors is the dimension of the space. $\text{rank}(A)$ is the dimension of $C(A)$.
   If $A^{n \times n}$ has linearly independent columns, A is invertible and the column vectors of A form a basis for $\mathbb{R}^n$. Dimension of $N(A) = n - \text{rank}(A)$.
Practice: Find the dimension of the vector space spanned by the vectors:

$$\begin{pmatrix} 1 \\ 1 \\ -2 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 0 \\ -4 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 3 \\ -3 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 0 \\ -2 \\ 0 \end{pmatrix}$$

And find a basis for that space.

Vector space of matrices

M = Set of all 3×3 matrices.

Practice: Show that the set of 2×3 matrices whose nullspace contains $\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$ is a vector subspace and find a basis for it. What about the set of those whose column space contains $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$?

Graph

If A is the incidence matrix and x is potential at nodes. Then Ax is the potential differences across nodes. Say C is the conductance per edge and f is the
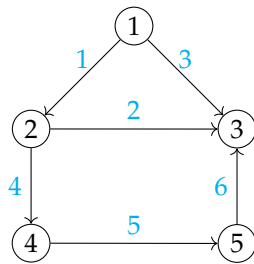
external input and output.

$$e = Ax$$
$$y = Ce$$
$$A^Ty = f$$

$$\boxed{A^TCAx = f}$$

- Find incidence matrix A

- $N(A), N(A^T) =$?

- $Trace(A^TA) =$?

Orthogonal vectors

The vectors x and y are orthogonal if $x^Ty = 0$.

Orthogonal subspaces

The subspaces S and T are orthogonal if every vector in S is orthogonal to every vector in T.

For $A^{m \times n}$, $C(A) \perp N(A^T)$ and $C(A^T) \perp N(A)$. And, $dimC(A) + dimN(A^T) = m$, $dimC(A^T) + dimN(A) = n$.

Solve $Ax = b$ when no exact solution

For $A^{m \times n}$ where $m > n$ or where too many equations, we may not have a solution. Solve below instead:
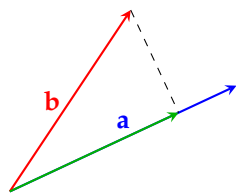
$$A^TAx = A^Tb$$

(A^TA) isn't always invertible. $N(A^TA) = N(A)$, $rank(A^TA) = rank(A)$. So, A^TA is invertible if A has linearly independent columns.

Practice: S is spanned by $(1, 2, 2, 3)$ and $(1, 3, 3, 2)$.

1. Find a basis for $S^{\perp}$

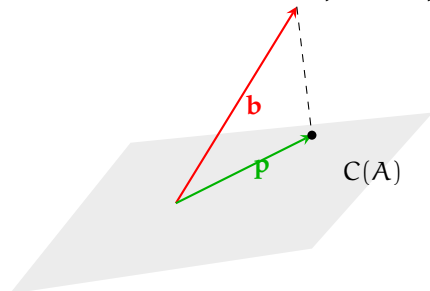2. Can every $v \in \mathbb{R}^4$ be written uniquely in terms of S and $S^{\perp}$?

Projection

Projection of the vector b onto the vector $a = a \frac{a^{\mathsf{T}} b}{a^{\mathsf{T}} a} = Pb$.

Here the projection matrix $P = \frac{1}{a^{\mathsf{T}} a} a a^{\mathsf{T}}$.

P is symmetric. $C(P)$ is the line through a. $\mathrm{rank}(P) = 1$. $P^2 = P$.

Why project?

Ax is in $C(A)$ and when $b \notin C(A)$ we have no solution for $Ax = b$. However, we can have solution to $A\hat{x} = p$ where p is b's projection onto $C(A)$.

Here $(b - p) \perp C(A)$ or $A^{\mathsf{T}}(b - p) = 0$.

$$A^{\mathsf{T}} b = A^{\mathsf{T}} p$$
$$A^{\mathsf{T}} b = A^{\mathsf{T}} A\hat{x}$$

As long as A has linearly independent columns, $A^{\mathsf{T}} A$ has inverse. In that case,

$$\hat{x} = \left(A^{\mathsf{T}} A\right)^{-1} A^{\mathsf{T}} b$$

$$p = A\hat{x} = A \left(A^{\mathsf{T}} A\right)^{-1} A^{\mathsf{T}} b$$

The projection matrix, $P = A \left(A^{\mathsf{T}} A\right)^{-1} A^{\mathsf{T}}$. P is symmetric. $P^2 = P$. On the other hand, $(I - P)$ projects b onto $N(A^{\mathsf{T}})$.

Least squares

$$c + d = 1$$
$$c + 2d = 2$$
$$c + 3d = 2$$

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$$

$$\hat{x} = \begin{pmatrix} \hat{c} \\ \hat{d} \end{pmatrix}$$

Practice: Find the orthogonal projection matrix onto the plane $x + y - z = 0$.

Least squares minimizes $\|e\|^2 = \|Ax - b\|^2$

Investigate: If we are looking at the data plane (x-y), the error least squares is minimizing looks like vertical distance to the fitted line. If we are looking at $C(A)$ and b, the error looks perpendicular to $C(A)$. Geometrically, why?

Investigate: What if we minimized the perpendicular distance to the fitted line instead of vertical distance to the fitted line as is being done here in least squares?

Practice: Find the quadratic equation through the origin that is a best fit for the points: $(1, 1)$, $(2, 5)$, and $(-1, -2)$.

Orthogonal matrices

Orthogonal matrix Q has orthonormal columns. $Q^\mathsf{T} Q = I$. If Q is square, $Q^\mathsf{T} = Q^{-1}$. Projection matrix, $P_Q = QQ^\mathsf{T}$.

Gram-Schmidt

Transform A with linearly independent columns into an orthogonal matrix Q. The first column is fine. For the second column, to make it perpendicular to the first column, subtract its projection onto the first column, the remainder is our new second column. For the third column, subtract its projections onto the first and the new second columns. So on. In the end, normalize these new columns to have Q.

For $\begin{pmatrix} \vdots & \vdots & \vdots \\ a & b & c \\ \vdots & \vdots & \vdots \end{pmatrix}$ with a, b, c linearly independent.

$\alpha = a$

$\beta = b - \dfrac{\alpha^\mathsf{T} b}{\alpha^\mathsf{T} \alpha} \alpha$

$\gamma = a - \dfrac{\alpha^\mathsf{T} c}{\alpha^\mathsf{T} \alpha} \alpha - \dfrac{\beta^\mathsf{T} c}{\beta^\mathsf{T} \beta} \beta$

$\alpha \perp \beta \perp \gamma$.

$$Q = \begin{pmatrix} \vdots & \vdots & \vdots \\ q_1 & q_2 & q_3 \\ \vdots & \vdots & \vdots \end{pmatrix} = \begin{pmatrix} \vdots & \vdots & \vdots \\ \frac{\alpha}{\|\alpha\|} & \frac{\beta}{\|\beta\|} & \frac{\gamma}{\|\gamma\|} \\ \vdots & \vdots & \vdots \end{pmatrix}$$

$A = QR$ where R is upper triangular.

Practice: Find q_1, q_2, q_3 (orthonormal) from a, b, c (columns of A). Then write A as QR.

$$A = \begin{pmatrix} 1 & 2 & 4 \\ 0 & 0 & 5 \\ 0 & 3 & 6 \end{pmatrix}$$

```
A = np.array(
    [
        [1, 2, 4],
        [0, 0, 5],
        [0, 3, 6]
    ]
)
a, b, c = A[:, 0], A[:, 1], A[:, 2]

project = lambda b, a: (a.T @ b)/(a.T @ a) * a
alpha = a
beta = b - project(b, alpha)
gamma = c - project(c, alpha) - project(c, beta)
Q = np.vstack(
    [
        alpha/la.norm(alpha),
        beta/la.norm(beta),
        gamma/la.norm(gamma)
    ]
)
R = Q.T @ A
```

Mean

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$\bar{x}_n = \bar{x}_{n-1} - \frac{1}{n} (x_n - \bar{x}_{n-1})$$

Notes:

- May lie outside of the sample.

- Outliers affect the mean.

- Adding c to each data point adds c to the mean.

- Multiplying each data point by c, multiplies the mean by c.

Variance

$$s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$s_n^2 = \frac{n-1}{n} s_{n-1}^2 + \frac{1}{n} (x_n - \bar{x}_{n-1}) (x_n - \bar{x}_n)$$

Notes:

- Average spread around the mean.

- Hard to interpret because in squared unit. So, the standard deviation, $\sqrt{s^2}$, is used.

- Adding c to each data point does not change the variance.

- Multiplying each data point by c, multiplies the variance by c^2.

Covariance Matrix

References

- Prof. Strang's lectures on Linear Algebra

- Coursera Math for ML Specialization