Linear algebra you need

208424 Lecture 2

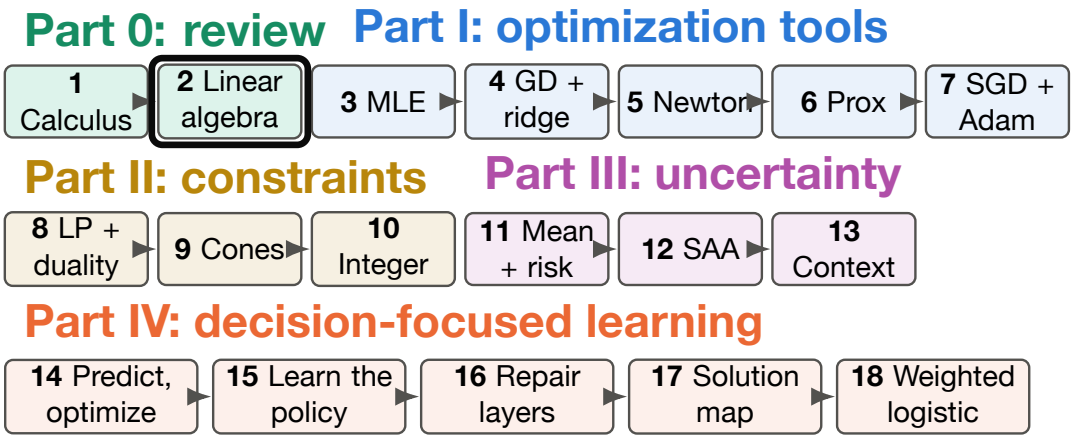
Data is a table

customer	income	debt ratio	default
A	30	0.2	no
B	55	0.6	yes
C	42	0.1	no

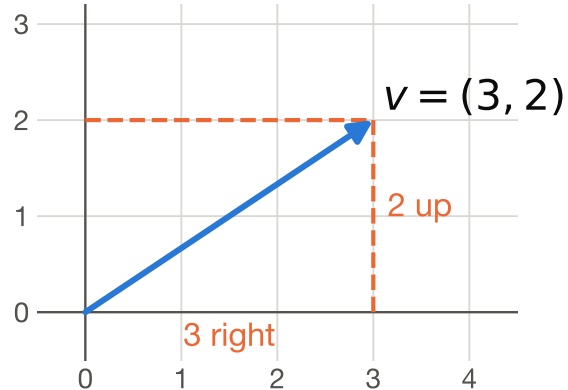
matrix X vector y

one row per customer · one column per feature

Course map


Part 0: review
Part I: optimization tools
1 Calculus
2 Linear algebra
3 MLE
4 GD + ridge
5 Newton
6 Prox
7 SGD + Adam
Part II: constraints
Part III: uncertainty
8 LP + duality
9 Cones
10 Integer
11 Mean + risk
12 SAA
13 Context
Part IV: decision-focused learning
14 Predict, optimize
15 Learn the policy
16 Repair layers
17 Solution map
18 Weighted logistic

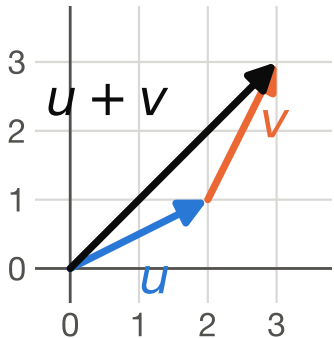
A vector is a list and an arrow



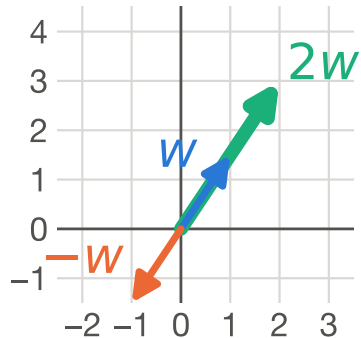
$$v = (3,\ 2)$$ = 3 right, 2 up

Add and scale

add: tip to tail

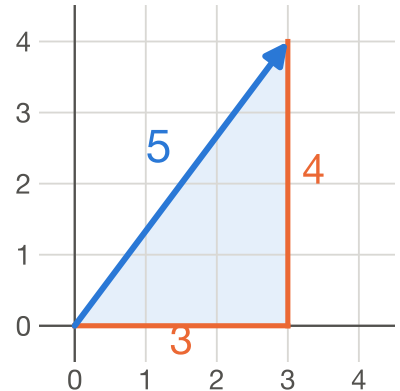


scale: stretch or flip



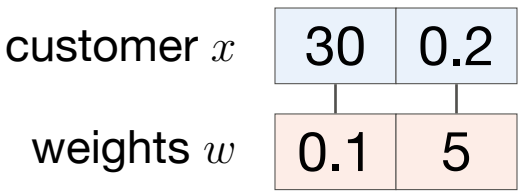
both work entry by entry

Length



$$\|v\| = \sqrt{3^2 + 4^2} = 5 \qquad \|w\|^2 = w_1^2 + \cdots + w_p^2$$

Dot product = weighted sum



$$30 \cdot 0.1 + 0.2 \cdot 5 = 4$$

$$x^\top w = x_1 w_1 + x_2 w_2 + \cdots + x_p w_p$$

multiply pairs, then add · $x^\top$: the transpose, slide 14

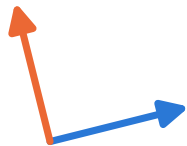
Dot product and angle

$u^\top v$ positive



same side

$u^\top v$ zero



perpendicular

$u^\top v$ negative



opposite side

the sign of $u^\top v$ shows the direction

Summary: vectors and the dot product

- Let $x = (x_1, \ldots, x_p)$ be the p features, w the weights.
- Add, scale entry by entry; length $\|w\| = \sqrt{w_1^2 + \cdots + w_p^2}$.
- We want: the score $x^\top w = x_1 w_1 + \cdots + x_p w_p$.
- Customer A: $x = (30,\ 0.2)$, $w = (0.1,\ 5)$, $x^\top w = 3 + 1 = 4$.
- Arrows u, v: the sign of $u^\top v$ shows the direction.

Checkpoint 1

$$u = (2,\ 1) \qquad v = (3,\ 4)$$

$$u^\top v = ? \qquad \|v\| = ?$$

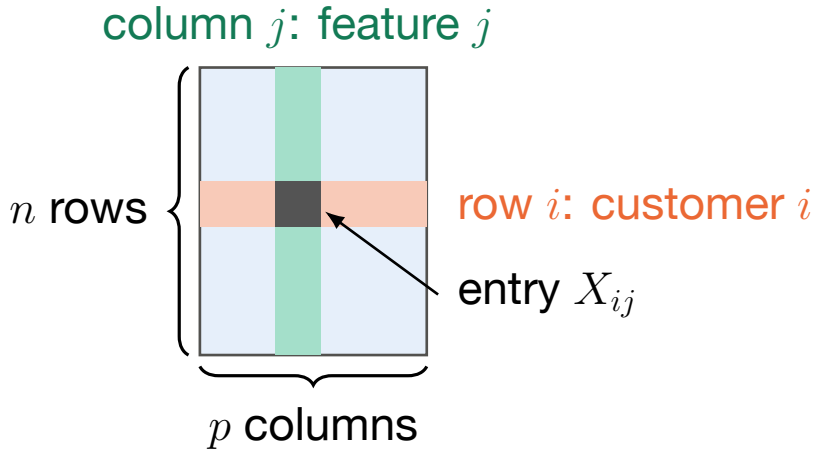
Checkpoint 1

$$u = (2,\ 1) \qquad v = (3,\ 4)$$

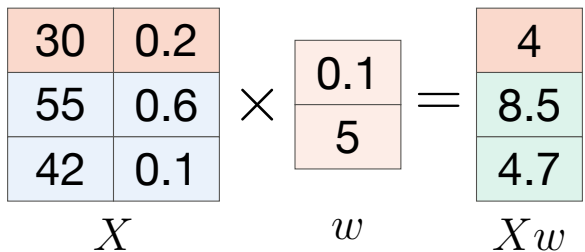
$$u^\top v = ? \qquad \|v\| = ?$$

$$2 \cdot 3 + 1 \cdot 4 = 10 \qquad \sqrt{9 + 16} = 5$$

A matrix is a table of numbers



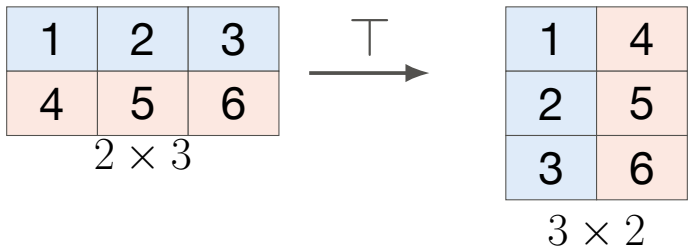
Xw: every prediction at once



entry i of Xw = row i of X dot w

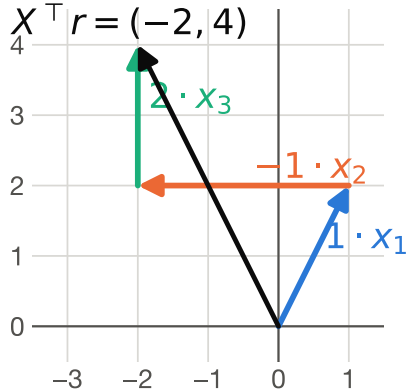
$$30 \cdot 0.1 + 0.2 \cdot 5 = 4$$

Transpose: rows become columns



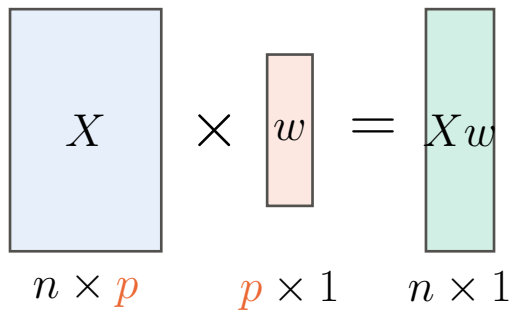
$$(X^\top)_{ij} = X_{ji}$$

$X^\top r$: a weighted sum of the rows



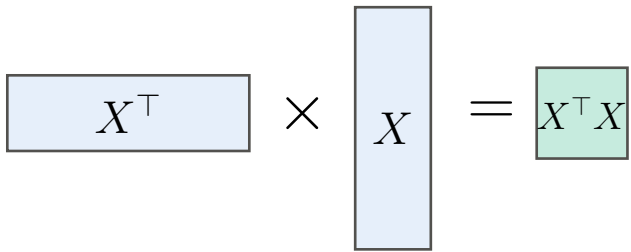
$X^\top r = \sum_i r_i \, x_i$ Lecture 3: r = prediction errors

When can we multiply?



the inner sizes must match: $(n \times p)(p \times q) = (n \times q)$

$X^\top X$: a small square matrix



1000 customers, 3 features $\Rightarrow$ $X^\top X$ is only 3×3

the rows of slide 15: $X^\top X = \begin{pmatrix} 10 & 2 \\ 2 & 5 \end{pmatrix}$, column dot column

Summary: matrices

- Let X be $n \times p$, row x_i = customer i, w the weights.
- We want: all n scores at once, Xw; entry i is $x_i^\top w$.
- Sizes: $(n \times p)(p \times q) = (n \times q)$ for any q: inner sizes match.
- $X^\top$ turns rows into columns, so $X^\top X$ is only $p \times p$.
- For $r = (r_1, \ldots, r_n)$: $X^\top r = \sum_i r_i \, x_i$, a sum of rows.

Checkpoint 2

X is 100×3. Size of Xw? Size of $X^\top X$?

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = ?$$

Checkpoint 2

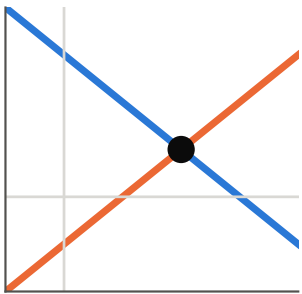
X is 100×3. Size of Xw? Size of $X^\top X$?

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = ?$$

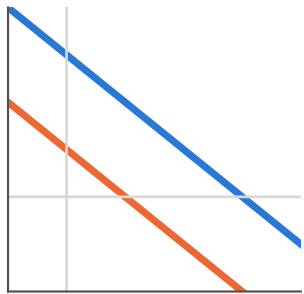
100×1 3×3 $\begin{pmatrix} 1+2 \\ 3+4 \end{pmatrix} = \begin{pmatrix} 3 \\ 7 \end{pmatrix}$

Two equations = two lines

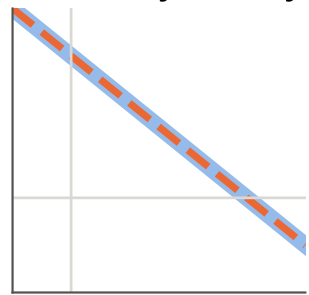
one solution



no solution

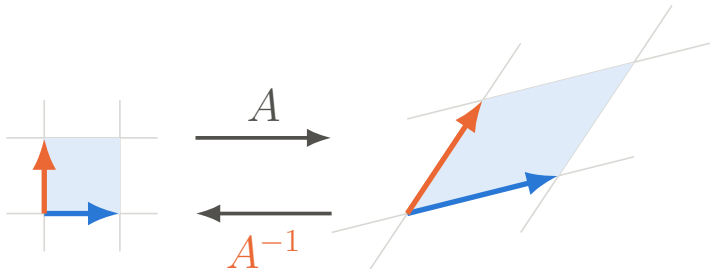


infinitely many



$Ax = b$: the solution is where the lines meet

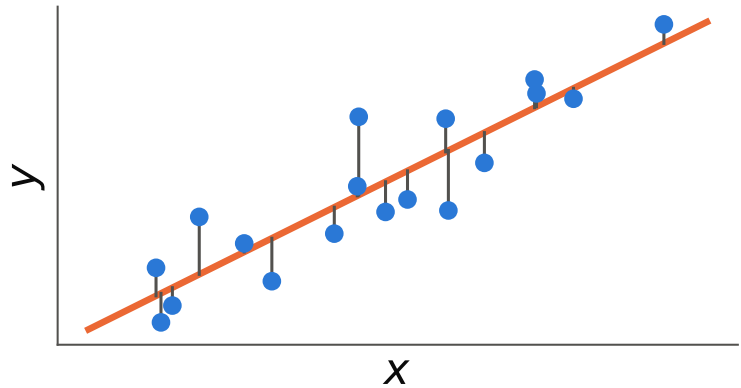
The inverse matrix A^{-1}



$A^{-1}A = I$ (identity)

$x = A^{-1}b$

Least squares: the normal equations



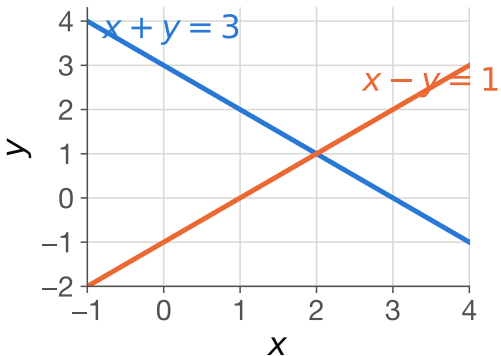
$$X^\top X\,\beta = X^\top y \;\Rightarrow\; \hat{\beta} = \left(X^\top X\right)^{-1} X^\top y$$

β: the weights · needs $X^\top X$ to have an inverse (slide 35)

Summary: solving $Ax = b$

- Let A = coefficients, x = unknowns, b = right-hand sides.
- We want: x with $Ax = b$, the point where the lines meet.
- The inverse: $A^{-1}A = I$ (identity), so $x = A^{-1}b$.
- Least squares, data X, y, weights β: $X^\top X\,\beta = X^\top y$, so $\hat{\beta} = (X^\top X)^{-1}X^\top y$.
- No A^{-1} if the lines are parallel (no x) or the same (many x).

Checkpoint 3

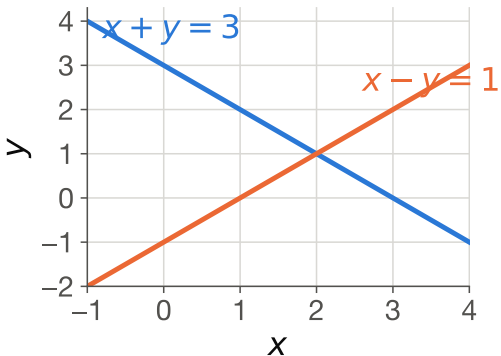


$$x + y = 3$$

$$x - y = 1$$

$$(x,\ y) = ?$$

Checkpoint 3



$$x + y = 3$$

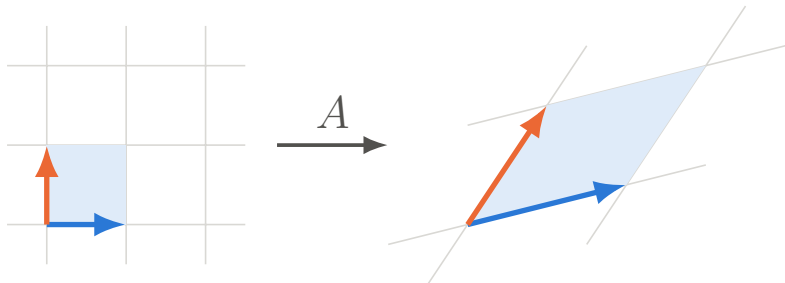
$$x - y = 1$$

$$(x,\ y) = ?$$

add: $2x = 4,\ \ x = 2$

then $y = 1$

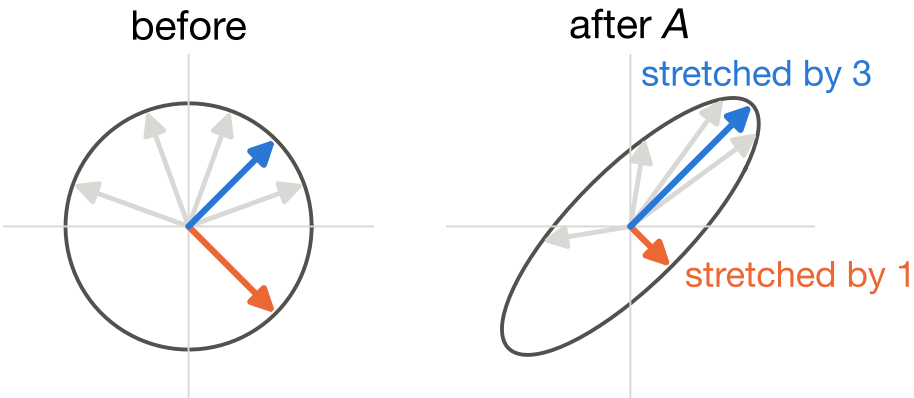
Multiplying every point by a matrix



$$A = \begin{pmatrix} 2 & 1 \\ 0.5 & 1.5 \end{pmatrix}$$

columns of A: $A(1,0)$ and $A(0,1)$

Eigenvectors and eigenvalues

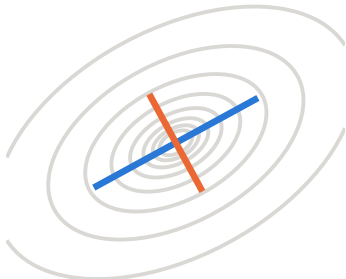


a new matrix A, rows $(2, 1)$ and $(1, 2)$

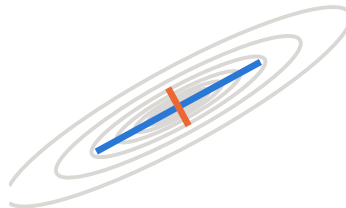
$Av = \lambda v$ eigenvalue λ = stretch factor

The condition number κ

λ = 3 and 1: κ = 3



λ = 25 and 1: κ = 25



$\kappa = \lambda_{\max}/\lambda_{\min}$ large κ: ill-conditioned (Lecture 4)

Summary: eigenvectors

- Let A be 2×2; its columns are $A(1,0)$ and $A(0,1)$.
- We want: arrows $v \neq 0$ that A only stretches or flips, $Av = \lambda v$.
- v is an eigenvector, λ its eigenvalue (the stretch factor).
- If A has rows $(2,1)$ and $(1,2)$: $\lambda = 3$ along $(1,1)$, $\lambda = 1$ along $(1,-1)$.
- All $\lambda > 0$: condition number $\kappa = \lambda_{\max}/\lambda_{\min}$; large κ: slow (Lecture 4).

The Hessian matrix

$$f(x, y) = x^2 + xy + 2y^2$$

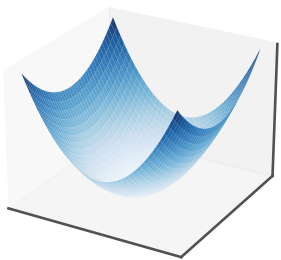
first partials: $\partial f/\partial x = 2x + y$, $\partial f/\partial y = x + 4y$

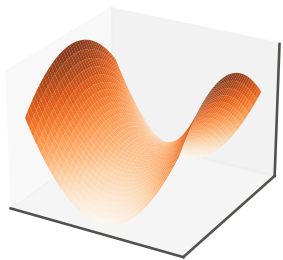
$$H = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x^2} = 2 & \dfrac{\partial^2 f}{\partial x\,\partial y} = 1 \\ \dfrac{\partial^2 f}{\partial y\,\partial x} = 1 & \dfrac{\partial^2 f}{\partial y^2} = 4 \end{bmatrix}$$

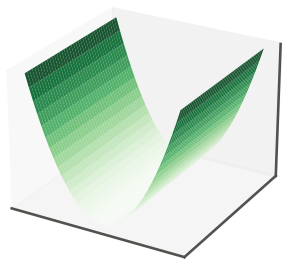
diagonal: curvature along an axis · off-diagonal: interaction

Quadratic forms: three shapes

bowl: +, + saddle: +, − valley: +, 0







$f(v) = v^\top A v$ the eigenvalue signs give the shape

The shape for a diagonal matrix

$$A = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}: \quad v^\top A v = a\, v_1^2 + b\, v_2^2$$

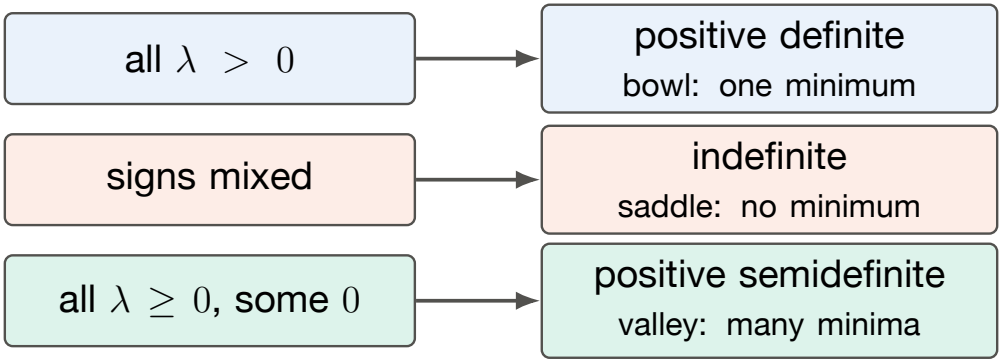
$a = 2,\ b = 1$	$2v_1^2 + v_2^2$: up in every direction	bowl
$a = 2,\ b = -1$	$2v_1^2 - v_2^2$: up along v_1, down along v_2	saddle
$a = 2,\ b = 0$	$2v_1^2$: flat along v_2	valley

diagonal A: eigenvalues a and b, since $A(1,0) = a(1,0)$

symmetric A: use its eigenvectors as axes (slide 28)

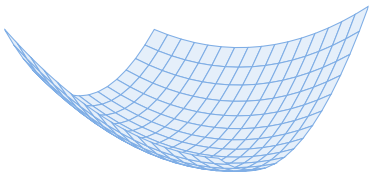
The eigenvalue test

for a symmetric A (every Hessian is symmetric)



all $\lambda \geq 0$: "positive semidefinite", written $A \succeq 0$

$X^\top X$ is positive semidefinite



$$v^\top X^\top X\, v$$

$$= (Xv)^\top (Xv) \quad ((Xv)^\top = v^\top X^\top)$$

$$= \|Xv\|^2$$

$$\geq 0 \quad \text{for every } v$$

least squares: never indefinite (never a saddle)
an eigenvalue 0 when one column of X is
a linear combination of the others

Summary: curvature with matrices

- Let H be the Hessian of f: its second derivatives.
- Near a stationary point x_0 ($\nabla f = 0$, Lecture 1), for a step v: $f \approx f(x_0) + \frac{1}{2} v^\top H v$.
- We want: the shape; let λ be the eigenvalues of H.
- All $\lambda > 0$: positive definite (bowl); mixed: indefinite (saddle); all $\lambda \geq 0$, some 0: positive semidefinite (valley).
- Any X: $v^\top X^\top X\, v = \|Xv\|^2 \geq 0$: $X^\top X \succeq 0$.

Checkpoint 4

$$A = \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix}$$

Is A positive definite, indefinite, or positive semidefinite?

Checkpoint 4

$$A = \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix}$$

Is A positive definite, indefinite,
or positive semidefinite?

eigenvalues 1 and -2: signs mixed $\Rightarrow$ indefinite (a

Lecture summary

$u^\top v$ positive



same side

$u^\top v$ zero



perpendicular

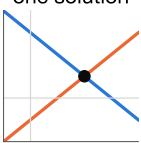
$u^\top v$ negative



opposite side

dot product = weighted sum

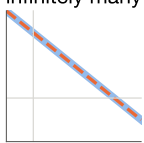
one solution



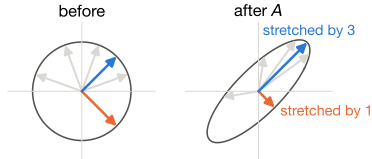
no solution



infinitely many

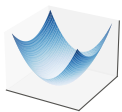


solve $Ax = b$

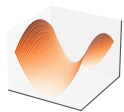


eigenvectors: $Av = \lambda v$

bowl: +, +



saddle: +, −

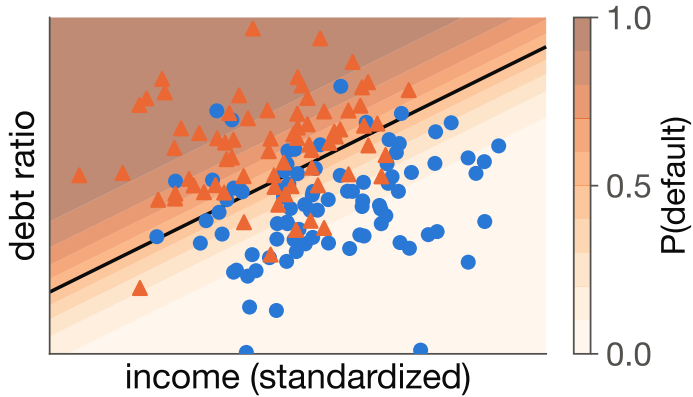


valley: +, 0



eigenvalues give the shape

Next lecture: estimation is optimization



$\sigma\left(x^\top w\right)$ choose w by minimizing a loss

Exercise

1. $u = (1, 2, 2)$, $v = (2, 0, 1)$: find $u^\top v$ and $\|u\|$.
2. X is 50×4: what is the size of $X^\top X$?
3. Solve $2x + y = 5$, $x - y = 1$.
4. $A = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}$: positive definite, indefinite, or positive semidefinite?