Review: **BILINEAR & QUADRATIC FORMS** LA 2 $\boxed{19|5|2026}$

Def: A **bilinear form** on a vector space V, over a field T:

(L1) $\forall x,y,z \in V,\ \forall t \in T:\ f(x+t\cdot y, z) = f(x,z) + t\cdot f(y,z)$

(L2) — " — : $f(x, y+t\cdot z) = f(x,y) + t\cdot f(x,z)$

- **symmetric**: $\forall x,y \in V,\ f(x,y) = f(y,x)$.

A **quadratic form** on a vector space V, over a field T, is a function $g: V \to T$ for which there exists a bilinear form f, $\forall x \in V,\ g(x) = f(x,x)$.

Def: The matrix of a bilinear form f w.r.t basis $B = \{b_1, \dots, b_n\}$ is the matrix A defined by $A_{ij} = f(b_i, b_j)$.

The matrix of a quadratic form g w.r.t the basis B is the matrix of a symmetric bilinear form f s.t. $g(u) = f(u,u)\ \forall u \in V$, if such f exists.

Lemma 1: If A is the matrix of a bilin. form f w.r.t. a basis $B = \{b_1, \dots, b_n\}$, then $\forall u,v \in V$: $f(u,v) = [u]_B^T A [v]_B$.

Lemma 2: If A is ...., then ${}_B[id]_C^T \cdot A \cdot [id]_{BC}$ is a matrix of f w.r.t. a basis C.

↖ transition matrix ...

Theorem: Let f be a symmetric bilinear form on VS V over T of characteristic $\neq 2$. Then there exists a basis B s.t. the matrix of f w.r.t B is diagonal.

Corollary (Sylvester's law of inertia).

For every quadratic form g on VS V of finite dimension over $\mathbb{R}$, there exists a basis B s.t. the matrix of g w.r.t. B is diagonal, only with entries $0, 1, -1$.

Moreover, for each such basis, the numbers of 1 and the number of -1 are invariant.

Proof: Let f be a symmetric bilin. form defining g, i.e., $\forall u \in V,\ g(u) = f(u,u)$.

By the previous Thm, there is a basis B of V s.t. the matrix of f w.r.t. B is diagonal - let D denote it.

Let S and D' are diagonal matrices defined as follows:

for i s.t. $D_{ii} = 0$: $S_{ii} = 1$, $D'_{ii} = 0$ i.e., $D'_{ii} = \operatorname{sgn} D_{ii}$

— " — $D_{ii} > 0$: $S_{ii} = \sqrt{D_{ii}}$, $D'_{ii} = 1$

— " — $D_{ii} < 0$: $S_{ii} = \sqrt{-D_{ii}}$, $D'_{ii} = -1$

Then $D = S^T \cdot D' \cdot S$, and as S is invertible,

$D' = (S^T)^{-1} \cdot D \cdot S^{-1}$ is diagonal, only with entries $0, 1, -1$.

By construction (c.f. lemma 2), D' is a matrix of the quadratic form g, with respect to some basis.

- invariance of the number of $+1$ and -1 on the diagonal:

Let D and D' be the diagonal matrices of g w.r.t. basis B and C. Then $[id]_{BC}^T \cdot D \cdot [id]_{BC} = D'$ (by lemma 2)

$\Rightarrow \operatorname{rank}(D) = \operatorname{rank}(D') \Rightarrow$ the same number of 0's

Assume, wlog, that $B = (b_1, \dots, b_n)$, $C = (c_1, \dots, c_n)$, and also assume, that both D and D' start with the positive elements $+1$ (permute the order in bases)

Let $p = \#1$ in D and $P = \mathcal{L}(\{b_1, \dots, b_p\})$ - linear span of $b_1, \dots, b_p$

$q = \#1$ in D' and $Q = \mathcal{L}(\{c_{q+1}, \dots, c_n\})$

For a contradiction, assume $q < p$.

$$D = \begin{array}{c} b_1 \dots b_p \\ \left[\begin{array}{ccc|c} 1 & & & \\ & \ddots & & 0 \\ & & 1 & \\ \hline & 0 & & \ddots \end{array}\right] \end{array} \qquad D' = \begin{array}{c} c_{q+1} \cdots c_n \\ \left[\begin{array}{ccc|c} 1 & & & \\ & \ddots & & 0 \\ & & 1 & \\ \hline & 0 & & \ddots \end{array}\right] \end{array}$$

Since $\dim(P) + \dim(Q) = p + n - q > n$, there exists

$v \in P \cap Q \Rightarrow v = \sum_{i=1}^{p} \beta_i b_i = \sum_{j=q+1}^{n} \gamma_j c_j$, for some β_i, γ_j.

$$\left.\begin{aligned} \text{Then } g(v) &= [v]_B^T \cdot D \cdot [v]_B = \sum_{i=1}^{p} \beta_i^2 > 0 \\ &= [v]_C^T \cdot D' \cdot [v]_C = \sum_{j=q+1}^{n} (-1) \cdot \gamma_j^2 \le 0 \end{aligned}\right\} \text{a contradiction}$$

$\Rightarrow q \geq p$. Symmetrically, we get $p \le q \Rightarrow p = q$ ∎

Example: let $g(x) = x^T A x$, for $A = \begin{pmatrix} 1 & 2 & 0 \\ 2 & -1 & 3 \\ 0 & 3 & -3 \end{pmatrix}$

Diagonalize:

$$A = \begin{pmatrix} 1 & & \\ 2 & -1 & \\ & 1 & 1 \end{pmatrix} \underbrace{\begin{pmatrix} 1 & & \\ & -3 & \\ & & 0 \end{pmatrix}}^{D} \begin{pmatrix} 1 & 2 & \\ & -1 & 1 \\ & & 1 \end{pmatrix}$$

Then for

$$S = \begin{pmatrix} 1 & & \\ & \sqrt{3} & \\ & & 1 \end{pmatrix}, \quad D' = \begin{pmatrix} 1 & & \\ & -1 & \\ & & 0 \end{pmatrix}$$ we have $D = S^T D' S$.

Corollary: In $\mathbb{R}^2$, there are six types of quadratic forms:

- classification by the number of $0, 1, -1$:

$$\underbrace{\overset{\text{"parabola3D"}}{\begin{pmatrix} 1 & \\ & 1 \end{pmatrix}}, \overset{\text{pringles}}{\begin{pmatrix} 1 & \\ & -1 \end{pmatrix}}, \overset{\text{"parabola3D" upside down}}{\begin{pmatrix} -1 & \\ & -1 \end{pmatrix}}}_{\text{paraboloid}}, \overset{\text{parabola 2D}}{\begin{pmatrix} 1 & \\ & 0 \end{pmatrix}}, \begin{pmatrix} -1 & \\ & 0 \end{pmatrix}, \overset{\text{plane}}{\begin{pmatrix} 0 & \\ & 0 \end{pmatrix}}$$

$$x_1^2 + x_2^2 \qquad x_1^2 - x_2^2 \qquad -x_1^2 - x_2^2 \qquad x_1^2 \qquad -x_1^2 \qquad 0$$

E.g. ↑

- satellite antenna
- mirror in reflectors
- positive definite matrices

---

in $\mathbb{R}^1$, three types only:

parabola		
(1)	(-1)	(0)
x_1^2	$-x_1^2$	0
		

RECAP of the year:

Winter semester: 1. Systems of linear equations, Gauss el.
2. Vector spaces, linear transformations
3. Matrices & computing with them

key notion: linear combination

Summer semester: 4. Inner products (1, 2, 3)
5. Determinant (1, 3)
6. Eigenvalues, eigenvectors (1, 2, 3, 5)
7. Positive definite matrices (3, 4, 5, 6)
8. Bilinear & quadratic forms (2, 4, 6, 7)

Matrix decomposition: QR, Cholesky, Sylvester

Important: matrix multiplication, EROs, Gauss el.

Recommendations for exam preparation

- Allocate enough time (a week)
- Review Linear algebra 1 - we build on it!
- Do not skip proofs! they help clarify what is it about
  → make it easier to rember theorem statements
- Think! Ask why?

GPT and LINEAR ALGEBRA

Generative Pre-traind Transformer

Game: Players take turns. On each turn, the player must say a word so that the sequence of words forms a sentence, or a sentence prefix. E.g. The cat is ...

Let V be the set of all words (tokens) (in GPT 3.5, $n = |V| \sim 200\,000$)
- a vocabulary

GPT 3: ~ 50000

An elementary task: simulate a player, i.e., given

INPUT: a sequence of words $w_1, w_2, \dots, w_k \in V$

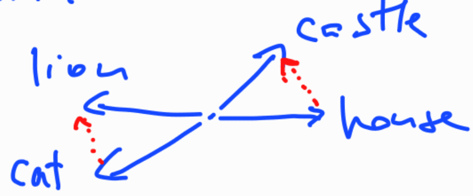
OUTPUT: a "meaningful" word $w_{k+1} \in V$

How to represent the words?

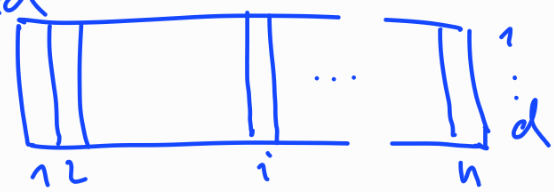
- not by a sequence of letters (the individual letters have no meaning)
- by vectors of a fixed lenght $d < n$ (GPT 3: 13000)

desire: similar words - similar vectors

e.g.



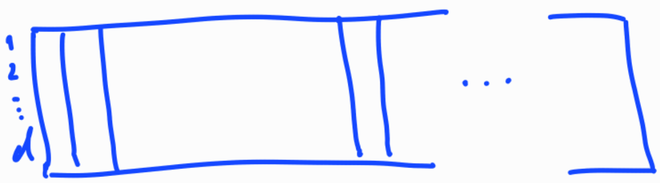
The mapping $e: V \to \mathbb{R}^d$ is precomputed (pre-traind), and represented by an embedding matrix W_e =



The words in V are ordered, the vector for the i-th word w is $e(w) = W_e \cdot e_i$

We will also need the reverse process - vector $\rightsquigarrow$ word:

an un-embedding matrix W_u^T =



given a vector $x \in \mathbb{R}^d$, normalization of $W_u x$ gives a probabilistic distribution on V

W_u is also pre-traind

How to represent the words in a sentence/text?

- idea 1: reflect the order of the words
- work with a context window of bounded lenght ℓ_{max} (GPT 3 ~ 1-3 tousands tokens, GPT 4 Turbo ~ 128 tousands, GPT 5.5 ~ 1 million)

$$W_p = \begin{array}{c|c} & 1\ 2\ \dots\ \ell_{max} \\ \begin{matrix}1\\2\\ \vdots \\ d\end{matrix} & \begin{bmatrix} \end{bmatrix} \end{array} \qquad p: \{1, \dots, \ell_{max}\} \to \mathbb{R}^d \qquad \text{pretrained}$$

$\Rightarrow$ initial vector representation of a word w_i: $x_i = e(w_i) + p(i)$

- idea 2: reflect the context (e.g., spring - several different meanings)

Done by attention mechanism: deals with two subproblems:

i) how much each of the words in the context window affects the word w_i?

intention: un-related words don't affect each other, related words clarify their meaning

how to know? by the inner product of vector representations x, x' of words w, w'.

un-related ~ orthogonal

ii) in which way (i.e., in which direction) each of the words $w_1, \dots, w_k$ in the context window affects the word w_i?

x_k - vector representation of w_k

x_k: $\langle x_k, x_1\rangle$ $\langle x_k, x_2\rangle$ $\dots$ $\langle x_k, x_k\rangle$ ... sequence of numbers $\in \mathbb{R}$

normalize s.t. all are non-negative, sum to one.

α_1 $\qquad \alpha_2$ $\qquad \dots$ $\qquad \alpha_k$ ~ probabil. distribution on $w_1, \dots, w_k$

instead of x_k, we use $\qquad x_k' = \sum_{i=1}^{k} \alpha_i \cdot x_i'$ (parallelizable)

What next?

apply the un-embedding on x_k': $y = W_u \cdot x_k' \in \mathbb{R}^n$

and normalize it: $z = \mathrm{Normal}(y) \in \mathbb{R}^n$. softmax

i.e., z is a probability distribution on V -

- pick the word w_{k+1} at random according to z.

Repeat.