INF1004 Linear Algebra Lecture 1: Vectors

Learning Goals I

- vectors of real numbers
- norms of vectors and their properties
- inner products, their interpretation and properties
- representing a vector as a linear combination
- vector spaces
- independent sets of vectors

Learning Goals II

- orthogonal sets of vectors
- projecting onto a vector, removing the direction of a vector
- creating an orthogonal set of vectors using the Gram-Schmid algorithm
- Matrix-vector and Matrix-Matrix multiplications

Part 1: Intro

Motivation for Linear Algebra

Examples for linear relationships in physics:

- Newtons first axiom

$$F = ma$$

is linear in the mass m and linear in the acceleration a.

- Einsteins energy-mass relationship for low speeds

$$E = mc^2$$

is linear in the mass m. This is the mass energy equivalence in special relativity theory for particles with a mass.

Motivation for Linear Algebra

Examples for linear relationships in physics:

- Linear relationships often hold as an approximation within in a certain range. Linear mapping as the first view on this world.

- Example: the force which a spring exerts when it is stretched by length x:

$$F = kx$$

- This does not hold when overstretched

- Einsteins "classic" energy mass formula also has its limits

Motivation for Linear Algebra

Examples for linear relationships in physics:

- Einsteins "classic" energy mass formula also has its limits. For high speeds one has to use:

$$E = \frac{mc^2}{\sqrt{1 - \frac{v^2}{c^2}}}$$

Motivation for Linear Algebra

Examples for linear relationships in physics:

- For mass-less photons the energy is given instead as:

$$E = hf$$

where $h = 6.626 * 10^{-34}\ J(Hz)^{-1}$ is Plancks constant and f the frequency (inverse of the wavelength) of the photon.

- Linear in the frequency f

Motivation for Linear Algebra

Examples for linear relationships in physics:

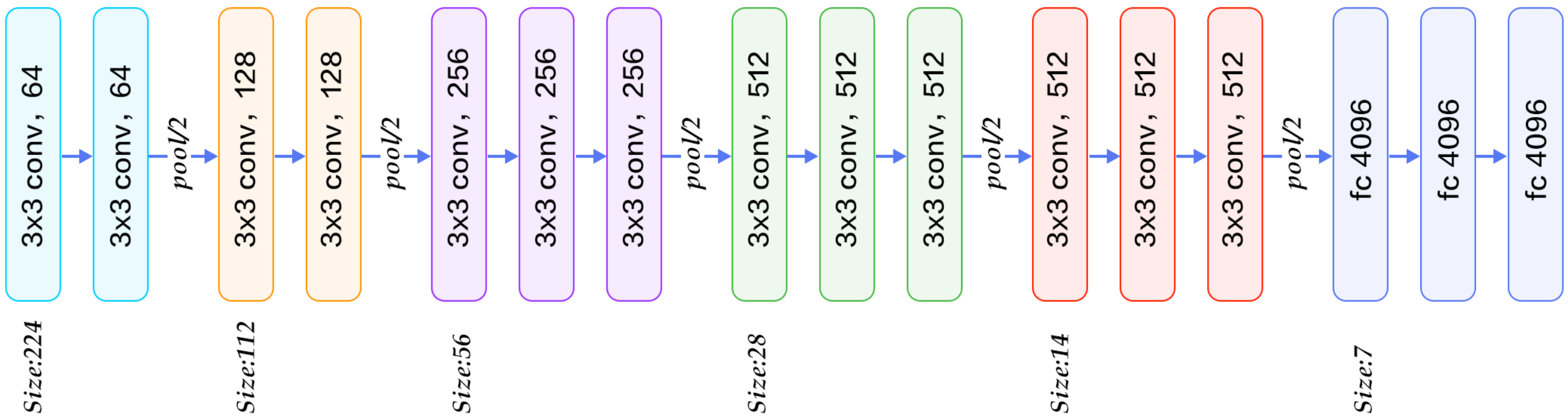
- The linear model as the simplest model for dependency on inputs
  - A score s as a weighted sum of causes x_0, x_1, x_2 ... provided the causes are on the same physical scale
  - and can be summed together. One can sum up three real numbers or vectors with weights.

$$s = w_0 x_0 + w_1 x_1 + w_2 x_2$$

Motivation for Linear Algebra

Examples for linear relationships in machine learning:

ML heavily relies on linear algebra: Neural nets are made of layers of linear operations, with non-linear activations on top.



Each block is a linear operation, implemented using matrix multiplication + activation function

Source https://www.kaggle.com/code/blurredmachine/vggnet-16-architecture-a-complete-guide/notebook

Motivation for Linear Algebra

Linear algebra is not essential for everything in life.

For a study with aspirations of doing any deeper (= better paid) work in Machine Learning or engineering it is needed.

Or you understand how to train better deepfakes!

https://www.dailymail.co.uk/news/article-6907549/Chinese-net-users-pay-just-90p-pretend-rich-kid-online-edited-photos-videos.html

Part 2: Vectors

Def: Vector of real numbers

Definition: A vector of real numbers is a sequence $v = (v_0, \ldots, v_{d-1})$

such that each component v_i is a real number. d is the dimensionality of the vector.

Examples:

$$v = (-1.1, 2.7, 3.5)$$
$$v = (0, 0, 1, 0, 0)$$
$$v = (2.4, 9.3)$$
$$v = (\cos(\alpha), \sin(\alpha))$$

Def: The space of d-dim vectors composed of real numbers

Definition: The space of d-dimensional vectors with real values $\mathbb{R}^d$ is the set of all sequences $(v_0, \ldots, v_{d-1})$ of length d such that each component v_i is a real number.

This vector space in literature is written as $\mathbb{R}^d$ because $\mathbb{R}$ denotes the real numbers.

Vectors in Physics

- A vector in a physical context can sometimes be interpreted as a measurement which has a magnitude and a direction.
- Examples:
  - Wind Speed with velocity of 3m/s and a direction of 227 degrees.
  - A pushing force in 2D of $(2, -3)$ Newton.

That is: 2 Newton along the first axis, and -3 Newton along the second axis – that is 3 newton against the direction of the second axis.

Its direction can be plotted:

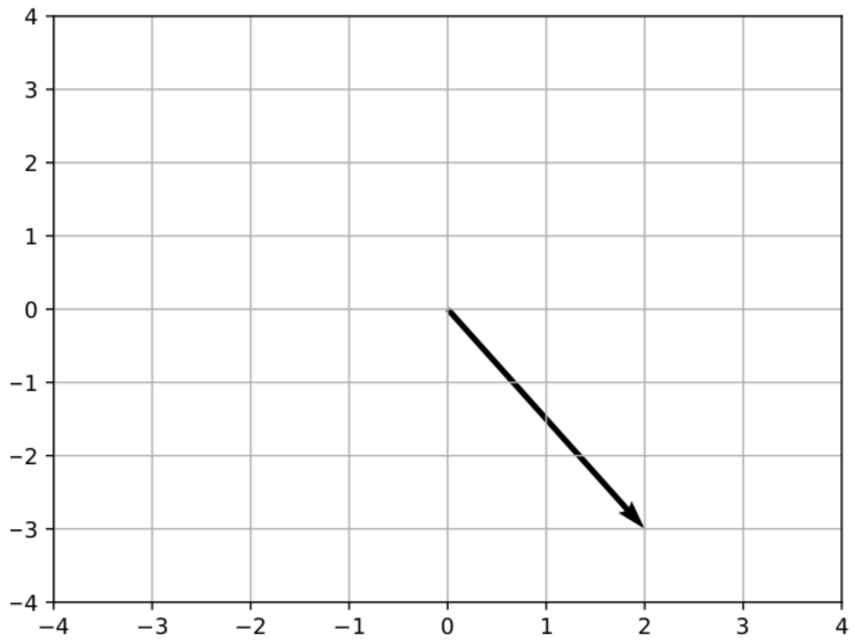
Vectors in Physics

- A vector in a physical context can sometimes be interpreted as a measurement which has a magnitude and a direction.
- A pushing force in 2D of (2, −3) Newton.

That is: 2 Newton along the first axis, and -3 Newton along the second axis – that is 3 newton against the direction of the second axis.

What is the magnitude and the angle in this example ?

Its direction can be plotted:



Vectors in ML

- Take any feature map, for example those obtained from a layer of a neural network. It is usually called a tensor. It has multiple indices. Flattening / Rolling out a tensor yields a vector .

Feature spaces do not have directions with geographical meanings (cf. previous section), but they still are composed of sequences of numerical values.

Part 3: Basic Properties of vectors

$v = \begin{bmatrix} 5 & 7 & 9 \end{bmatrix}$ in pytorch: $v.shape = (1, 3)$ or $v.shape = (3)$

$v = \begin{bmatrix} 5 \\ 7 \\ 9 \end{bmatrix} = (5, 7, 9)^\top$ in pytorch: $v.shape = (3, 1)$ or $v.shape = (3)$

Row and column vector shapes

Can be written as column or row vector:

$v = \begin{bmatrix} 5 & 7 & 9 \end{bmatrix}$ in pytorch: $v.shape = (1, 3)$ or $v.shape = (3)$

$v = \begin{bmatrix} 5 \\ 7 \\ 9 \end{bmatrix} = (5, 7, 9)^\top$ in pytorch: $v.shape = (3, 1)$ or $v.shape = (3)$

Vector Addition: component-wise

$$v = (v_0, v_1, v_2),\ w = (w_0, w_1, w_2)$$

$$v + w = (v_0 + w_0, v_1 + w_1, v_2 + w_2)$$

Example:

$$(1, 2, 3) + (2, -1, 3) = (3, 1, 6)$$

Vector Scaling: component-wise

$$v = (v_0, v_1, v_2)$$
$$av = (av_0, av_1, av_2)$$
$$5(2, -1, 3) = (10, -5, 15)$$

Part 4: Vector norms

Vector Norms: Euclidean Norm

This is also known as ℓ2 -norm

$$\|v\|_2 = \left(\sum_{k=1}^{d} v_k^2\right)^{1/2}$$

This corresponds to the usual idea of classic Euclidean lengths. In physics the usual magnitude of a vector.

Example: $\|(-1,3,-2)\|_2 = \sqrt{(-1)^2 + 3^2 + (-2)^2} = \sqrt{1+9+4} = \sqrt{14} \approx ?$

One can define other norms.

Vector Norms: 1-norm, p-norms

The 1-norm:

$$\|v\|_1 = \sum_{k=1}^{d} |v_k|$$

p-norms in General:

$$\|v\|_p = (\sum_{k=1}^{d} |v_k|^p)^{1/p}$$

Note that the Euclidean and the 1-norm are special cases of this for p = 2 and p = 1

Vector Norms: maximum-(infinity)-norm

The maximum-norm:

$$\|v\|_\infty = \max_{k=0,\ldots,d-1} |v_k|$$

It is a limit of p-norms for $p \to \infty$

Out of exams: Use cases for other norms

- Setup: One wants to obtain a linear prediction of y based on x

$$y = f_w(x) = \sum_{k=0}^{d-1} w_k x_k$$

Inputs: pairs of a vector and its corresponding desired output

$$(x^{\{i\}}, y^{\{i\}}) = ((x_0^{\{i\}}, \ldots, x_{d-1}^{\{i\}}), y^{\{i\}})$$

Out of exams: Use cases for other norms

- Standard, unregularized, linear regression would try to find the best weights

$w = (w_0, \ldots, w_{d-1})$ based on a set of training samples minimizing the squared difference between prediction $f(x^{\{i\}})$ and desired value $y^{\{i\}}$:

find w such that $\frac{1}{n}\sum_{i=1}^{n}(y^{\{i\}} - \sum_{k} w_k x_k^{\{i\}})^2$

$= \frac{1}{n}\|Y - Xw\|_2^2$ will be minimal

Out of exams: Use cases for other norms

Linear regression with 1-norm penalty is called LASSO-regression.

- LASSO adds a 1-norm penalty on the weights.

find w such that $\frac{1}{n}\sum_{i=1}^{n}(y^{\{i\}} - \sum_k w_k x_k^{\{i\}})^2 + \lambda\|w\|_1$

$= \frac{1}{n}\|Y - Xw\|_2^2 + \lambda\|w\|_1$ will be minimal

- Effect: one can ensure that the weights are sparse, that is that we have $w_k = 0$ for many weight vector components.

Vector Norms: properties of any norm

$$||0|| = 0$$

$$||\lambda v|| = |\lambda| ||v||$$

$$||v + u|| \leq ||v|| + ||u||$$

a norm induces a distance measure $d(\cdot, \cdot)$ via $d(u, v) = ||u - v||$

Norms are always positive or zero.

General: Norms encode a notion of length of a vector.

Unit length Vectors - for the Euclidean norm

Definition: v is unit length vector with respect to the euclidean norm if

$$\|v\|_2 = 1$$

Turn a vector into one with unit length:

$$v \neq 0 \Rightarrow \frac{v}{\|v\|_2}$$

The result has unit length. (This works also with every other norm).

Part 5: Inner Product

Inner Product of Vectors (the canonical one)

Let be $u \in \mathbb{R}^d,\, v \in \mathbb{R}^d$. Definition: Then the inner product between u and v is defined as:

$$u \cdot v = \sum_{k=0}^{d-1} u_k v_k$$

Note: The output of the inner product is always a real or complex number, never a vector.

Example: $$(3, -1, -2) \cdot (1, -2, 1.5) = 3 + 2 - 3 = 2$$

Inner Product of Vectors

Let be $u \in \mathbb{R}^d, v \in \mathbb{R}^d$. Definition: Then the inner product between u and v is defined as:

$$u \cdot v = \sum_{k=0}^{d-1} u_k v_k$$

One can define other inner products!
We do not discuss them in this lecture.

Interpretation of the Inner Product

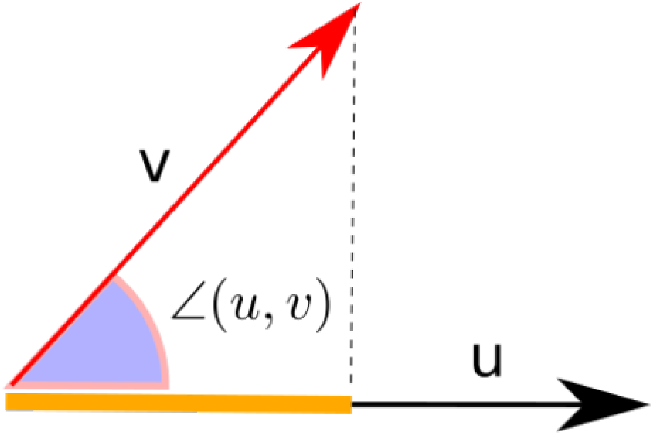
It holds for the inner product defined above that:

$$u \cdot v = ||u||_2 ||v||_2 \cos(\angle(u, v))$$

It is the product of the euclidean length of u, of v and the cosine of the angle between these two vectors.

See the cosine here: https://en.wikipedia.org/wiki/Sine_and_cosine#/media/File:Sine_cosine_one_period.svg

Interpretation of the Inner Product



has length equal to

$\|v\|_2 \cos(\angle(u, v))$

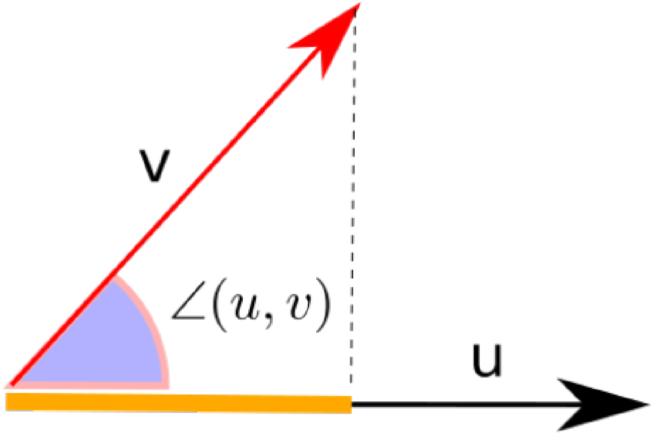
$$\frac{u}{\|u\|_2} \cdot \frac{v}{\|v\|_2} = \cos(\angle(u, v))$$

is a similarity measure between two vectors.

The cosine angle gets larger if the angle between the two vector decreases:

- it is −1 if u = −c v, c<0
- it is 0 for orthogonal vectors
- it is 1 is u = c v , c>0

Interpretation of the Inner Product



has length equal to
$\|v\|_2 \cos(\angle(u, v))$

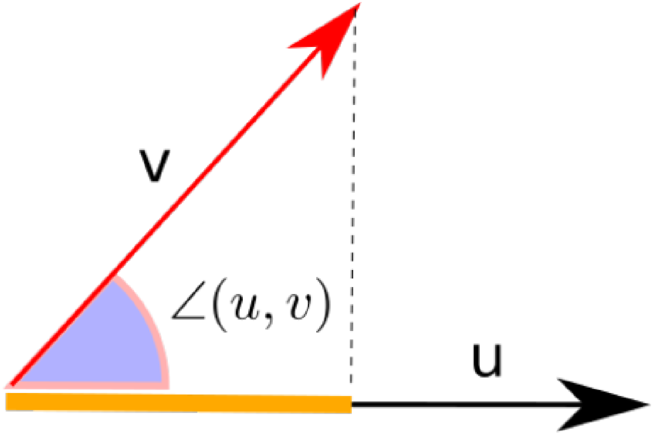
$$\frac{u}{\|u\|_2} \cdot \frac{v}{\|v\|_2} = \cos(\angle(u, v))$$

is a similarity measure between two vectors.

The cosine angle gets larger if the angle between the two vector decreases:

- for angles between 90 and 270 degrees (in radians the interval (π/2, 3/2π)), the cosine and thus the inner product is negative

Interpretation of the Inner Product



has length equal to $\|v\|_2 \cos(\angle(u,v))$

$$\frac{u}{\|u\|_2} \cdot \frac{v}{\|v\|_2} = \cos(\angle(u,v))$$

is a similarity measure between two vectors.

The cosine angle gets larger if the angle between the two vector decreases:

- for angles between 270 and 90 degrees (in radians the intervals (0, π/2) ∪ (3/2π, 2π)), the cosine and thus the inner product is positive

Properties of the Inner Product

1. It is symmetric in its two arguments: $u \cdot v = v \cdot u$

2. It is a bilinear mapping, that is it is linear in each of its two input arguments:

Let a_1, a_2 be real numbers, $u^{\{1\}}, u^{\{2\}}, v$ be vectors. Then:

Linearity in the first argument: $(a_1 u^{\{1\}} + a_2 u^{\{2\}}) \cdot v = a_1(u^{\{1\}} \cdot v) + a_2(u^{\{2\}} \cdot v)$

Linearity in the 2nd argument: $v \cdot (a_1 u^{\{1\}} + a_2 u^{\{2\}}) = a_1(v \cdot u^{\{1\}}) + a_2(v \cdot u^{\{2\}})$

Meaning: linear combinations and inner product can be swapped

Properties of the Inner Product

2. It is a bilinear mapping, that is it is linear in each of its two input arguments:

Let a_1, a_2 be real numbers, $u^{\{1\}}, u^{\{2\}}, v$ be vectors. Then:

$$(a_1 u^{\{1\}} + a_2 u^{\{2\}}) \cdot v = a_1 (u^{\{1\}} \cdot v) + a_2 (u^{\{2\}} \cdot v)$$

$$v \cdot (a_1 u^{\{1\}} + a_2 u^{\{2\}}) = a_1 (v \cdot u^{\{1\}}) + a_2 (v \cdot u^{\{2\}})$$

Inner product of a linear combination = linear combination of inner products

Properties of the Inner Product

2. It is a bilinear mapping, that is it is linear in each of its two input arguments:

- Warning: It is not linear as a function of the two inputs u,v viewed as one concatenated vector (u,v)!

$$L((u,v)) = u \cdot v$$

$$L(a(u,v)) = L((au,av)) = (au) \cdot (av) \qquad = a^2 u \cdot v \neq aL((u,v))$$

This is the difference between linear and bilinear: bilinear = linear in each of the two arguments. Being a linear mapping would mean: it is linear as a function of a single argument (the concatenation of both vectors).

Properties of the Inner Product

3. Every inner product defines a norm $\|u\|_{(\cdot)} = \sqrt{u \cdot u}$

4. Inner products with the zero vector: $u \cdot 0 = 0$

Note that the 0 here are different types on the left and on the right hand side

Motivation for the Inner Product

- matrix-vector and matrix-matrix multiplications compute a set of inner products
- fully connected layers in Neural nets: https://pytorch.org/docs/stable/generated/torch.nn.Linear.html (see also torch.functional) compute a vector-matrix multiplication

$$y = xA^\top + b$$

Which is a set of inner products

Motivation for the Inner Product

- Convolution Layers: https://github.com/vdumoulin/conv_arithmetic

They compute inner products between the convolution kernel and a (sliding) window in the input features.
(You do not need to understand the mechanism of convolution in detail. It is not part of graded knowledge.)

- Conclusion: these layers compute inner products, and thus similarities (weighted by the norms of both vectors) between the input features x and the trainable weights w.

Part 6: Representability as a linear combination, Vector Spaces, Linear Independence

Representability as a linear combination

Motivation: This will be useful to measure quantities along coordinate axes, to express vectors y as a function of a few selected vectors.

Representability of a vector y as a linear combination of others:

A vector y can be represented as a linear combination over a set of vectors $\{v^{\{0\}}, \ldots, v^{\{d-1\}}\}$, if there exist real numbers $a_0, \ldots, a_{d-1}$ such that

$$y = \sum_{k=0}^{d-1} a_k v^{\{k\}}.$$

Representability as a linear combination

Examples:

- (1,0) and (0,1) can represent all vectors (a_0, a_1)
- (1,1) and (0,1) can also represent all vectors (a_0, a_1)

- Each element tv of the line $\{tv, t \in \mathbb{R}\}$ is a linear combination over the vector v
- (1,0,0) and (0,1,0) cannot represent a vector (a_0, a_1, a_2) if $a_2 \neq 0$.

- {(1,0,0), (0,1,0), (0,0,1)} can represent any vector (a_0, a_1, a_2)

The general definition of a Vector space

A set of objects V is a vector space if three properties are met:

1. we know how to multiply an object $v \in V$ with a real number, that is we can define a multiplication operation $a * v$, where $v \in V$ is an object from the set, and a is a real number
2. we know how to add two objects $u \in V, v \in V$, that is we can define an addition operation $u + v$
3. the set V is closed under the above two operations, that is if $u \in V, v \in V$, then their multiplications with real numbers a_1, a_2 and sums of vectors are also contained within the set: $u \in V, v \in V \Rightarrow a_1 u + a_2 v \in V$

The general definition of a Vector space

The short version:
A set of "stuff" is a vector space V if

- We can multiply stuff with real (or complex) numbers
- We can add two stuffs together
- Linear combination of two or more stuffs stays within the set of "stuff"

- Examples for such other stuff are polynomials and continuous functions.
- Counterexample: all vectors with norm below 1.

Vector space spanned by a set of vectors

The vector space spanned by a set of vectors $w^{\{0\}}, \ldots, w^{\{n-1\}}$ is the set of all their linear combinations using all possible real numbers $a_0, \ldots, a_{n-1} \in \mathbb{R}$

$$V = \{w \quad \text{such that} \quad w = \sum_{r=0}^{n-1} a_r w^{\{r\}}, a_0, \ldots, a_{n-1} \in \mathbb{R}\}$$

The dimension of this set is the largest number of independent vectors which we can obtain from $w^{\{0\}}, \ldots, w^{\{n-1\}}$

Break ?

Part 7: (Linearly) independent sets of vectors

(Linearly) independent set of vectors

A set of vectors $\{v^{\{0\}}, \ldots, v^{\{d-1\}}\}$ is independent (in the sense of linear algebra), if no vector $v^{\{i\}}$ from this set can be represented by the vectors from the set $\{v^{\{0\}}, \ldots, v^{\{d-1\}}\} \setminus \{v^{\{i\}}\}$ (that is from the set without using $v^{\{i\}}$ itself).

That is for any $v^{\{i\}}$ from the set $\{v^{\{0\}}, \ldots, v^{\{d-1\}}\}$, there exists **no solution** of shape

$$v^{\{i\}} = \sum_{k=0, k \neq i}^{d-1} a_k v^{\{k\}}.$$

(Linearly) independent set of vectors

Next step: show why independent sets are useful.

(Linearly) independent set of vectors

If $\{v^{\{0\}}, \ldots, v^{\{d-1\}}\}$ is an independent set with a fixed ordering of vectors (sets are unordered!!), and we have a decomposition of a vector

$y = \sum_{k=0}^{d-1} a_k v^{\{k\}}$, then the coefficients $\{a_0, \ldots, a_{d-1}\}$ are unique.

That is, if there exists another decomposition of y: $y = \sum_{k=0}^{d-1} b_k v^{\{k\}}$ with respect to the same set with the same ordering of vectors, then it must hold $a_k = b_k$ for all $k = 0, \ldots, d-1$.

Proof by contradiction. We assume we have two decompositions of y, and they are different in one index r. Then we show that we can represent $v^{\{r\}}$ as a linear combination of the other vectors.

Basis in a finite dimensional vector space and dimension of a vector space

A set of vectors $\{v^{\{0\}}, \ldots, v^{\{d-1\}}\}$ is a basis of a vector space V, if the set is independent, and any vector $y \in \mathcal{V}$ can be represented as a linear combination of the set. The number of vectors in this set, d, is the dimensionality of the vector space.

Note: There is no clash with the definition of independence.
$v^{\{i\}}$ from the set can be represented as a linear combination of the set, by using only itself.

One possible Basis of $\mathbb{R}^d$

Basis of $\mathbb{R}^d$

The set

$$
\begin{aligned}
\{e^{\{0\}}, \ldots, e^{\{d-1\}}\} &\text{ with} \\
e^{\{0\}} &= \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \end{bmatrix}^\top \\
e^{\{1\}} &= \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \end{bmatrix}^\top \\
e^{\{1\}} &= \begin{bmatrix} 0 & 0 & 1 & \cdots & 0 \end{bmatrix}^\top \\
e^{\{k\}} &= \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \begin{matrix} \\ \\ \\ \leftarrow \text{at pos. } k \\ \\ \\ \end{matrix} \\
e^{\{d-1\}} &= \begin{bmatrix} 0 & 0 & 0 & \cdots & 1 \end{bmatrix}^\top
\end{aligned}
$$

is one possible basis of $\mathbb{R}^d$. Its elements are called one-hot vectors.

One possible Basis of $\mathbb{R}^d$

One hot vectors have a simple formula, below for the k-th one hot vector:

$$e_i^{\{k\}} = \begin{cases} 1 \text{ if } i = k \\ 0 \text{ if } i \neq k \end{cases}$$

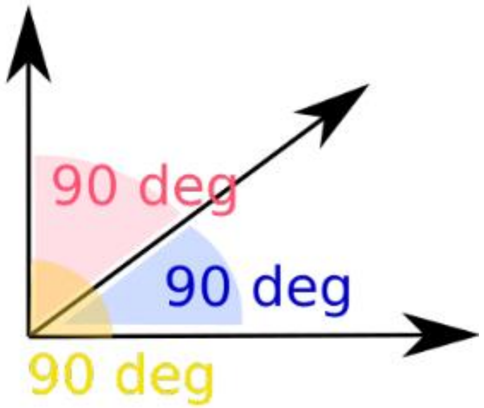
Break ?

Part 8: Orthogonal sets of vectors

Orthogonal sets of vectors

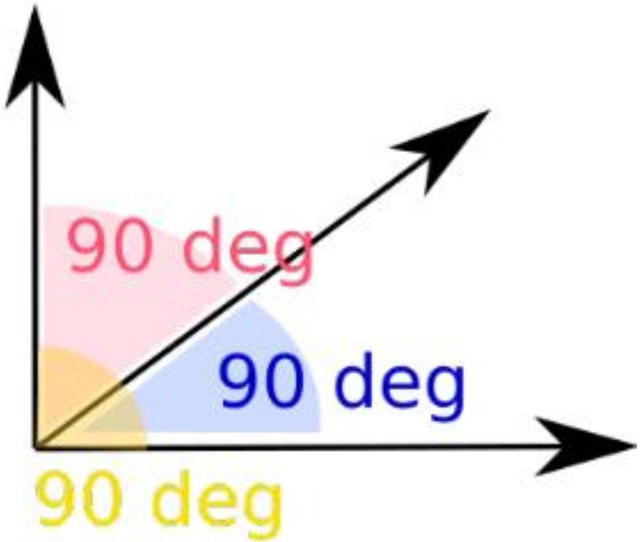
$\{v^{\{0\}}, \ldots, v^{\{d-1\}}\}$ is an orthogonal set of vectors if all $v^{\{i\}} \neq 0$ and $i \neq k \Rightarrow v^{\{i\}} \cdot v^{\{k\}} = 0$

Obviously such a set has angles of 90 degrees between each pair of different vectors. What is it good for ?



They are suitable as vectors defining the axes in a coordinate system!

Orthogonal sets of vectors



- You can think of a set of orthogonal sets as vectors defining the axes of a coordinate system, no matter how high dimensional the space is.
- The point of having axes with 90 degrees between each other is to measure … for every vector x in the space ... the component of x along one chosen axis independent of the other axes.
- It is about measurement and representation of vectors.

Orthogonal sets of vectors

What is the value of having an orthogonal set ?

- For an orthogonal set, we can compute the coefficients a_k of a linear combination very easily.

- Suppose we want to represent y as a linear combination of the $v^{\{k\}}$ and find out the corresponding coefficients a_k

$$y = \sum_{k=0}^{d-1} a_k v^{\{k\}}$$

compute inner product of both sides of the eq. with $v^{\{0\}}$

$$y \cdot v_0 = \left(\sum_{k=0}^{d-1} a_k v^{\{k\}} \right) \cdot v^{\{0\}}$$

$$= \sum_{k=0}^{d-1} a_k \; v^{\{k\}} \cdot v^{\{0\}}$$

$$= a_0 \; v^{\{0\}} \cdot v^{\{0\}} + 0$$

$$\frac{y \cdot v^{\{0\}}}{v^{\{0\}} \cdot v^{\{0\}}} = a_0$$

Orthogonal sets of vectors

The same holds if we take the inner product with any $v^{\{r\}}$

$$y = \sum_{k=0}^{d-1} a_k v^{\{k\}}$$

Take the inner product with $v^{\{r\}}$ on both sides

$$\Rightarrow \frac{y \cdot v^{\{r\}}}{v^{\{r\}} \cdot v^{\{r\}}} = \frac{y \cdot v^{\{r\}}}{\|v^{\{r\}}\|_2^2} = a_r$$

Orthogonal sets of vectors

If $\{v^{\{0\}}, \ldots, v^{\{d-1\}}\}$ is an Orthogonal set of vectors and y can be linearly represented with respect to this set, that is $y = \sum_{k=0}^{d-1} a_k v^{\{k\}},$

Then we have: $\frac{y \cdot v^{\{r\}}}{v^{\{r\}} \cdot v^{\{r\}}} = \frac{y \cdot v^{\{r\}}}{\|v^{\{r\}}\|_2^2} = a_r$

Orthogonal sets of vectors

A real number

This has a clean interpretation:

If $\{v^{\{0\}}, \ldots, v^{\{d-1\}}\}$ is an Orthogonal set of vectors, then $\frac{y \cdot v^{\{r\}}}{\|v^{\{r\}}\|_2^2}$

Is the amount of vector $v^{\{r\}}$ which is present in vector y.

Orthogonal sets are linearly independent sets.

Proof by contradiction. We assume that an orthogonal set would be not independent, and use above coefficient formula which uses inner products to show the contradiction.

Orthogonal sets of vectors

In short, orthogonal sets are useful:

- they are an independent set, thus coefficients of a linear representation are unique
- when trying to represent a vector we can get the coefficients in a simple way
- as we will see below, there is an algorithm how to obtain an orthogonal set (with some possible zero vectors as undesired radioactive waste)

Part 9: Projecting a vector on another vector. Removing a vector from another.

Projecting a vector on another vector. Removing a vector from another.

1. Projecting a vector x onto a vector v:

$$x_{\|v} = \frac{x \cdot v}{v \cdot v} v = \left(x \cdot \frac{v}{\|v\|_2} \right) \frac{v}{\|v\|_2}$$

This makes use of the "clean interpretation" of the coefficient

Projecting a vector on another vector. Removing a vector from another.

2. Removing a vector v from a vector x:

$$x_{\perp v} = x - x_{\| v} = x - \frac{x \cdot v}{v \cdot v} v = x - \left(x \cdot \frac{v}{\|v\|_2} \right) \frac{v}{\|v\|_2}$$

You can derive this when you know how to project onto a vector v.

This means: $x_{\perp v}$ has no component in direction of vector v, that is $x_{\perp v} \cdot v = 0$.

We can show that $x_{\perp v}$ it is indeed orthogonal to v

$$x_{\perp v} \cdot v = x \cdot v - x_{\| v} \cdot v = x \cdot v - \frac{x \cdot v}{v \cdot v}(v \cdot v) = x \cdot v - x \cdot v = 0$$

Removing a vector from another.

Avoid this source of mistake:

Above works if you remove the direction of a single vector from x.

However (!!!), if you subtract two vectors $v^{\{0\}}$ and $v^{\{1\}}$ from x, and these two vectors share a component in common $(v^{\{0\}} \cdot v^{\{1\}} \neq 0)$, then you can reintroduce a component back by mistake.

Here an extreme example, where we use for convenience $v^{\{0\}} = v^{\{1\}} = v$

$$x_{\text{bad } v^{\{0\}}, v^{\{1\}}} = x - x_{\| v^{\{0\}}} - x_{\| v^{\{1\}}} = x - x_{\| v} - x_{\| v} = x - 2x_{\| v}$$

$$x_{\text{bad } v^{\{0\}}, v^{\{1\}}} \cdot v = 0 - 1x_{\| v} \cdot v = -x \cdot v$$

Removing a vector from another.

Avoid this source of mistake:

Above works if you remove the direction of a single vector from x.

However (!!!), if you subtract two vectors $v^{\{0\}}$ and $v^{\{1\}}$ from x, and these two vectors share a component in common $(v^{\{0\}} \cdot v^{\{1\}} \neq 0)$, then you can reintroduce a component back by mistake.

Good news: one can avoid this mistake by applying this iteratively, that is removing $v^{\{1\}}$ from $x_{\perp v^{\{0\}}}$ (and not from x itself as wrongly done above)!

Part 10: Constructing an orthogonal set using inner products

This is also known as Gram-Schmidt orthogonalization.

Constructing an orthogonal set using inner products

It has two properties:

1. The input is a set of vectors. The output is a set of orthogonal vectors and some of the vectors in the output can be zero vectors .
2. The vector space spanned by the first i output vectors is the same as the vector space spanned by the first i input vectors.

Constructing an orthogonal set using inner products

Run a for loop over the number of vectors

- Step 0: $$\widetilde{v}^{\{0\}} = v^{\{0\}}$$

- Step 1: Remove from v {1} the direction of v~{0} , as much as there is present:

$$\widetilde{v}^{\{1\}} = v^{\{1\}} - v^{\{1\}}{}_{\|\widetilde{v}^{\{0\}}}$$

Constructing an orthogonal set using inner products

Run a for loop over the number of vectors

- Step 2: Remove from v {2} the directions of v~{0} and v~{1} , as much as there is present:

$$\begin{aligned}\widetilde{v}^{\{2\}} &= v^{\{2\}} - v^{\{2\}}{}_{\|\widetilde{v}^{\{0\}}} - v^{\{2\}}{}_{\|\widetilde{v}^{\{1\}}} \\ &= v^{\{2\}} - \sum_{k=0}^{1} v^{\{2\}}{}_{\|\widetilde{v}^{\{k\}}}\end{aligned}$$

- Continue the same in all other steps for the next vectors

Constructing an orthogonal set using inner products

Orthogonalization (numerically not so stable!)

- Algorithm Input: non-zero vectors $v^{\{0\}}, \ldots, v^{\{d-1\}}$
- Step 0: Initialize:
$$\tilde{v}^{\{0\}} = v^{\{0\}}$$
- run a for loop over $0, \ldots, d-1$
- step r (in a `for`-loop from $r = 1$ to $r = d-1$): have got so far: $\tilde{v}^{\{0\}}, \ldots, \tilde{v}^{\{r-1\}}$
$$\begin{aligned}\tilde{v}^{\{r\}} &= v^{\{r\}} - \sum_{k=0}^{r-1} \frac{v^{\{r\}} \cdot \tilde{v}^{\{k\}}}{\tilde{v}^{\{k\}} \cdot \tilde{v}^{\{k\}}} \tilde{v}^{\{k\}} \\ &= v^{\{r\}} - \sum_{k=0}^{r-1} v^{\{r\}}{}_{\|\tilde{v}^{\{k\}}}\end{aligned}$$

  at each step: Remove zero vectors $\tilde{v}^{\{k\}} = 0$ (you cannot divide by $\|\tilde{v}^{\{k\}}\|_2^2$ for a zero vector).
- Return the non-zero vectors from $\tilde{v}^{\{0\}}, \ldots, \tilde{v}^{\{d-1\}}$ as orthogonal set

Constructing an orthogonal set using inner products

disadvantage: Gram-Schmidt can result in numerical problems when implemented on a computer due to subtractive cancellation. One may get vectors which are

- not-zero but only almost zero when they should be,
- and will have non-zero inner products to the other result vectors.

It is suitable, however, to understand how to construct orthogonal bases for subspaces.

A numerically more stable version is given on the next slide

Constructing an orthogonal set using inner products

A numerically more stable version is:

Modified Gram Schmidt (better numerical stability)

- for k in range(0,d): $\widetilde{v}^{\{k\}} = v^{\{k\}}$ init
- for k in range(0,d):
  - $\widetilde{z}^{\{k\}} = \frac{\widetilde{v}^{\{k\}}}{\|\widetilde{v}^{\{k\}}\|_2}$
  - for m in range(k+1,d): $\widetilde{v}^{\{m\}} = \widetilde{v}^{\{m\}} - (\widetilde{v}^{\{m\}} \cdot \widetilde{z}^{\{k\}})\widetilde{z}^{\{k\}}$

The difference is: when one obtained a normalized $\widetilde{z}^{\{k\}}$, then one removes its component from all "future" $\widetilde{v}^{\{l\}}$ with indices l > k. Then one moves forward by incrementing k = k + 1.

A property of Gram-Schmid

Any vector u which can represented as a linear combination of the vectors $v^{\{0\}}, \ldots, v^{\{i\}}$ up to index i, can also be represented by the output $\widetilde{v}^{\{0\}}, \ldots, \widetilde{v}^{\{i\}}$ of the Gram-Schmidt process up to the same index i. This holds for every index i.

Gram-Schmidt not only allows to create an orthogonal set, but the output has the same ability for representing vectors (as linear combination) as the input to the algorithm -- at each step of the algorithm.

A property of Gram-Schmid

A proof of it is in the extended pdf.

Part 11: Matrix Multiplications

Matrix Multiplications

Given a matrix of shape (n, d) and a vector of dimensionality d, the multiplication Ax of them – with x being on the right hand side is defined as a vector of length n such that

$$Ax = \begin{bmatrix} a_{0,0} & \cdots & a_{0,d-1} \\ a_{1,0} & \cdots & a_{1,d-1} \\ \vdots & \ddots & \vdots \\ a_{n-1,0} & \cdots & a_{n-1,d-1} \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ \cdots \\ x_{d-1} \end{bmatrix}$$

$$= \begin{bmatrix} [a_{0,0} \ \cdots \ a_{0,d-1}] \cdot x \\ [a_{1,0} \ \cdots \ a_{1,d-1}] \cdot x \\ \vdots \\ [a_{n-1,0} \ \cdots \ a_{n-1,d-1}] \cdot x \end{bmatrix}$$

The k-th component of this vector is given as:

$$\begin{aligned} (Ax)_k &= A[k,:] \cdot x \\ &= (a_{k,0} \ldots a_{k,d-1}) \cdot x \\ &= \sum_{r=0}^{d-1} a_{k,r} x_r \end{aligned}$$

Matrix Multiplications

Given a matrix of shape (n, d) and a vector of dimensionality n, the multiplication $x^\top A$ of them – with $x^\top$ being on the left hand side is defined as a vector of length d such that

$$x^\top A = [x_0, x_1, \ldots, x_{n-1}] \begin{bmatrix} a_{0,0} & \cdots & a_{0,d-1} \\ a_{1,0} & \cdots & a_{1,d-1} \\ \vdots & \ddots & \vdots \\ a_{n-1,0} & \cdots & a_{n-1,d-1} \end{bmatrix}$$

$$= \left[x \cdot \begin{bmatrix} a_{0,0} \\ \vdots \\ a_{n-1,0} \end{bmatrix} \quad x \cdot \begin{bmatrix} a_{0,1} \\ \vdots \\ a_{n-1,1} \end{bmatrix} \quad \cdots \quad x \cdot \begin{bmatrix} a_{0,d-1} \\ \vdots \\ a_{n-1,d-1} \end{bmatrix} \right]$$

The k-th component of this vector is given as:

$$\begin{aligned} (x^\top A)_k &= x \cdot A[:, k] \\ &= (a_{0,k} \ldots a_{n-1,k}) \cdot x \\ &= \sum_{r=0}^{n-1} x_r a_{r,k} \end{aligned}$$

Matrix Multiplications (with a matrix)

Given a matrix A of shape (n, d) and a matrix B of shape (d, f), their multiplication AB is defined as a matrix of shape (n, f) and its component $(AB)_{i,k}$ at row i and column k is given as

$$(AB)_{i,k} = A_{(i,:)} \cdot B_{(:,k)} = \sum_{r=1}^{d} A_{i,r} B_{rk}$$

$$AB = \begin{bmatrix} A_{(0,:)} \cdot B_{(:,0)} & A_{(0,:)} \cdot B_{(:,1)} & A_{(0,:)} \cdot B_{(:,2)} & \cdots & A_{(0,:)} \cdot B_{(:,f-1)} \\ A_{(1,:)} \cdot B_{(:,0)} & A_{(1,:)} \cdot B_{(:,1)} & A_{(1,:)} \cdot B_{(:,2)} & \cdots & A_{(1,:)} \cdot B_{(:,f-1)} \\ A_{(2,:)} \cdot B_{(:,0)} & A_{(2,:)} \cdot B_{(:,1)} & A_{(2,:)} \cdot B_{(:,2)} & \cdots & A_{(2,:)} \cdot B_{(:,f-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_{(n-1,:)} \cdot B_{(:,0)} & A_{(n-1,:)} \cdot B_{(:,1)} & A_{(n-1,:)} \cdot B_{(:,2)} & \cdots & A_{(n-1,:)} \cdot B_{(:,f-1)} \end{bmatrix}$$

Matrix Multiplications (with a matrix)

Therefore a matrix- matrix multiplication is a matrix consisting of inner products:

$$AB = \begin{bmatrix} A_{(0,:)} \cdot B_{(:,0)} & A_{(0,:)} \cdot B_{(:,1)} & A_{(0,:)} \cdot B_{(:,2)} & \cdots & A_{(0,:)} \cdot B_{(:,f-1)} \\ A_{(1,:)} \cdot B_{(:,0)} & A_{(1,:)} \cdot B_{(:,1)} & A_{(1,:)} \cdot B_{(:,2)} & \cdots & A_{(1,:)} \cdot B_{(:,f-1)} \\ A_{(2,:)} \cdot B_{(:,0)} & A_{(2,:)} \cdot B_{(:,1)} & A_{(2,:)} \cdot B_{(:,2)} & \cdots & A_{(2,:)} \cdot B_{(:,f-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_{(n-1,:)} \cdot B_{(:,0)} & A_{(n-1,:)} \cdot B_{(:,1)} & A_{(n-1,:)} \cdot B_{(:,2)} & \cdots & A_{(n-1,:)} \cdot B_{(:,f-1)} \end{bmatrix}$$

Matrix Multiplications (with a matrix)

Memorizing

- Matrix-Matrix multiplication results in a matrix, if shapes are permissible:

  $(n, d)(d, f) \to (n, f)$.

- The component $(AB)_{i,k}$ at row i and column k is given as the inner product between row i of the left matrix and column k of the right matrix.

  That also tells you which axes must match in dimensionality: left matrix - the number of columns $=$ the dimensionality of the second axis. right matrix - the number of rows $=$ the dimensionality of the first axis.

Inner Product as matrix-vector multiplication

Consequence:

inner products as matrix vector multiplication

The inner product $x \cdot y$ of two column-shaped vectors can be written in matrix-vector multiplication notation as

$$x \cdot y = \left[x_0, x_1, \ldots, x_{d-1}\right] \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{d-1} \end{bmatrix} = x^\top y$$

where $x^\top$ is the transpose of a vector or matrix x.

The transpose was used here to convert the column-shaped vector into a row-shaped vector.