# Complete Solutions to Lecture 1 Assignments

Minh Nguyen, Long Le, Giang Hoang, and Hung Ngo

Last update: 18:41, Sunday 12th April, 2020

**Caution**: Personal use only.
Do not distribute, to save the fun solving the problems for next classes!

## Contents



## 1 From Lecture Slides

### P 1.3

> Why is $x_{\text{new}} = (1 - \alpha)x_{\text{old}} + \alpha x_{\text{update}}$ called exponentially weighted?

#### Notation

For convenience, we consider $x_{\text{new}}$ and $x_{\text{old}}$ consecutive elements of a sequence $(x_i)$. We also denote $x_{\text{update}}$ by a corresponding element of $(g_i)$.

#### Explanation for the term *exponentially weighted*

An explicit formula for $x_t$ can be worked out as

$$x_t = \alpha(g_t + (1-\alpha)g_{t-1} + (1-\alpha)^2 g_{t-2} + \dots) + (1-\alpha)^t x_0.$$

Now take a look at the first term. The sum inside the brackets is weighted with $((1-a)^i)_{i=1}^t$. These weights decay *exponentially*!

### Notes

We recommend you watch Prof. Andrew Ng's explanation ([part 1], [part 2]) on exponentially weighted averages.

## P 1.4

> Prove that these are vector spaces:
>
> - The real $n$-space $\mathbb{R}^n$ (also called the **standard** Euclidean space).
> - Matrices $\mathbb{R}^{m\times n}$ and multi-dimensional arrays (data "tensors") $\mathbb{R}^{m\times n\times \dots q}$.
> - $C[a, b]$–the set of all continuous functions on the closed interval $[a, b]$.
> - The space of all polynomials of one variable $x$ whose degree is at most $n$, $P_n(\mathbb{R}) := \{\sum_{i=0}^n a_i x^i\}$.

In the following proofs, we concern mainly about the closure of addition and scalar multiplication. The other axioms are are skipped here since they are relatively simple to verify if we can define our operations clearly.

#### $\mathbb{R}^{m\times n\times \dots q}$

Note that this case is a generalization of $\mathbb{R}^n$ and $\mathbb{R}^{m\times n}$. As addition and scalar multiplication are element-wise for vectors, matrices, and tensors, the problem reduces to the case of $\mathbb{R}$. This is indeed a vector space: real sums and real products are real! We can thus conclude the same for $\mathbb{R}^{m\times n\times \dots q}$.

#### $C[a, b]$

Multiplying functions with scalars and adding them together have nothing to do with their domain. Therefore, continuity is preserved and $C[a, b]$ is a vector space.

$P_n(\mathbb{R})$

Scaling $P_n(\mathbb{R})$ polynomials and adding them by no means give raise to terms of degree $n+1$ and higher. These operations are hence closed and $P_n(\mathbb{R})$ is a vector space as desired.

## P 1.5

> Prove that these are *not* vector spaces:
>
> - The space of positive real axis.
> - Unit vectors.
> - Latitude and longitude.
> - Monomials $\{x^k\}$.

In the following proofs, we will point out features of the given spaces which do not follow vector space axioms.

### The space of positive real axis

This is not a vector space since it lacks additive inverses (which are supposed to be negative numbers).

### Unit vectors

This is not a vector space as the sum of two unit vectors is not a unit vector.

### Latitude and longitude

We can treat this problem with latitudes and longitudes either defined as individual spaces or combined together into a 2-dimensional space.

Firstly, addition with latitudes is undefined in its nature. But let's say we agree to set north latitudes as positive numbers and their south counterpart as negative numbers. Even so, addition in this space is not closed: many of them will result in sums that are greater than $90°$. As such, latitudes do not form a vector space.

The same argument works for longitude, with minor modifications–east means positive and west means negative. If longitudes constituted a vector space, we would get weird longitudes of over $180°$!

The case of combined coordinates with latitudes and longitudes has become clearer after our previous arguments. A coordinate vector $[x\ y]$ where $x$ and $y$ are respectively out of the intervals $[-90°, 90°]$ and $[-180°, 180°]$ surely doesn't lie in our desired vector space.

#### Monomials

This not a vector space since addition for monomials is not closed: $x^m + x^n$ is no longer a monomial.

### Geometric intuition

> Use geometric intuitions to prove the following:

#### P 1.6

> A **line** *through* $v$ in *direction* of $u$: $L_1 = \{w \in V : w = v+tu,\ \forall t \in \mathbb{R},\ u, v \in V\}$.

By geometric intuition, any point $w$ on the described $L_1$ can be considered as $v$ translated in direction of $u$, hence $w = v + tu$ for some scalar $t$. Now let $t$ vary $t \in \mathbb{R}$ we have $L_1$.

#### P 1.7

> A **line through 2 points** $u \neq v$: $L_2 = \{w \in V : w = (1-t)v + tu,\ \forall t \in \mathbb{R},\ u, v \in V\}$.

Apply P1.6 with direction $u' = u - v$ then rearrange the terms.

#### P 1.8

> A space $S$ is **flat** if any line between any 2 points of $S$ is contained in it $\Rightarrow$ *All vector spaces are flat*!

Any line $L$ between any 2 vectors in $V$ is just a linear combination of them, as in P1.6-7, hence $L \subset V$ by definition (of vector space). Hence $V$ is flat.

**P 1.9**

> A line through $u$ *and* the **zero vector** $0_V$ (think of it as the "origin" of the abstract space): $L_0 = \{w \in V : w = tu,\ \forall t \in \mathbb{R},\ u \in V\}$. *Translating* $L_2$ by $p \neq 0_V$ we get $L_0' \parallel L_2$ ($L_0'$ passes through $0_V$).

Apply P1.6 with $v = 0_V$ we have $L_0$. Translate $L_2$ in P1.7 by $-v$ we get $L_0'$ passing through the origin and the point $(u - v)$.

**P 1.10**

> A line **segment** between 2 points $u$ & $v$: $L_3 = \{w \in V : w = (1-t)v + tu,\ \forall t \in [0,1],\ u, v \in V\}$.

Apply P1.7 with the translation $t(u-v)$ strictly in direction from $v$ to $u$, i.e., $t \geq 0$, and the translation is bounded within the segment, i.e., $t(u-v) \leq u - v \Leftrightarrow t \leq 1$.

**P 1.11**

> What happens if we let $v$ in the lines above *move* along another line which is affine/convex combination of 2 different vectors $u_2, v_2$? Repeat for $v_2$? Then prove that with $\sum_i \alpha_i = 1$ the linear combination $\alpha_1 x_1 + \cdots + \alpha_n x_n$ is an affine/convex combination (the latter also requires $\alpha_i > 0\ \forall i$).

From P1.7, consider $L = \{w \in V : w = (1-t_1)v_1 + t_1 u_1,\ \forall t_1 \in \mathbb{R},\ u_1, v_1 \in V\}$. Now fix $u_1$ and let $v_1$ moves on another line $L' = \{v_1 \in V : v_1 = (1-t)v_2 + tu_2$. This creates a *plane* through 3 points $P \ni w = (1-t_1)(1-t)v_2 + t_1u_1 + t(1-t_1)u_2$. Let $t_2 = t(1-t_1)$ we have $P = \{w \in V : w = (1-t_1-t_2)v_2 + t_1u_1 + t_2u_2,\ \forall t_1, t_2 \in \mathbb{R},\ u_1, u_2, v_2 \in V\}$. Repeat the same reasoning for $v_2$ and so on, we have the affine combination of $n$ vectors. Similarly for convex combination.

## P 1.12

> What are the standard basis (and associated coordinate vectors) of $\mathbb{R}^d$, $\mathbb{R}^{m\times n}$, and $P_n(\mathbb{R})$?

$\mathbb{R}^d$

The standard basis of $\mathbb{R}^d$ is the tuple $(x_i)_{i=1}^d$ where vector $x_i \in \mathbb{R}^d$ has $1$ for entry $i$ and $0$ for all other. The associated coordinate vectors are vectors in $\mathbb{R}^d$ themselves.

$\mathbb{R}^{m\times n}$

The standard basis of $\mathbb{R}^{m\times n}$ is the tuple $(X_i)_{i=1}^{mn}$ where vector $X_i \in \mathbb{R}^{m\times n}$ has $1$ for entry $([i/n], i - n[i/n])$ and $0$ for all other. The coordinate vectors are vectors in $\mathbb{R}^{mn}$ (notice there's no $\times$ between $m$ and $n$).

$P_n(\mathbb{R})$

The standard basis of $P_n(\mathbb{R})$ is the monomial tuple $(x^i)_{i=0}^n$. Its coordinate vectors are vectors in $\mathbb{R}^{n+1}$.

## P 1.14

> Consider vector space $\mathbb{R}^3$. Let's choose a basis $B$ with 3 vectors:
>
> $$e_1 = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0), e_2 = (-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0), e_3 = (0, 0, 1)$$
>
> What are the coordinates in this basis $B$ of a vector in $\mathbb{R}^3$?

**Notation**

Let's say we have a vector $x = [x_1\ x_2\ x_3]^\top \in \mathbb{R}^3$. Our task is to represent $x$ as

$$x = y_1 e_1 + y_2 e_2 + y_3 e_3, \tag{14.1}$$

i.e., with coordinate vector $[y_1\ y_2\ y_3]^\top$ in basis $B$. This eventually reduces to solving for $y_1$, $y_2$, and $y_3$ in terms of $x_1$, $x_2$, and $x_3$.

**Finding $y_1$, $y_2$, and $y_3$**

The desired coefficients in (14.1) is exactly the solutions to the following system of equations:

$$\begin{cases} x_1 &= \dfrac{y_1 - y_2}{\sqrt{2}} \\ x_2 &= \dfrac{y_1 + y_2}{\sqrt{2}} \\ x_3 &= y_3 \end{cases}$$

As the third equation already defined $y_3$ directly, we only need to manipulate the first two to finally get

$$\begin{cases} y_1 &= \dfrac{x_1 + x_2}{\sqrt{2}} \\ y_2 &= \dfrac{x_2 - x_1}{\sqrt{2}} \\ y_3 &= x_3 \end{cases}$$

**Geometric view**

You can think of $B$ as the standard basis gone through the $\frac{\pi}{4}$rad rotation around the $z$-axis. Therefore, coordinates in this new basis have to be rotated back the same $\frac{\pi}{4}$rad from the standard coordinates.

# 2 Coding Exercises

### C 1.1

> Write a `genData(N,d,'dist')` function which randomly generates $N$ length-$d$ vectors following distribution `dist`.

To be updated.

### C 1.2

> Write a `avgSample(sample,w)` function which computes the mean of `sample` given a weight vector $w$. Extend this to other central tendency measures.

#### Suggestive key function

See **code 1** for reference. Note that `preprocess()` and `weight_preprocess()` functions are needed to standardize our data and weights.

```python
import numpy as np
import pandas as pd

# ... many lines after

def central_tendency(data, tendency, **arg):

  '''
  Return specified central tendency of
  given data.

  Parameters:
  - data: array_like. The data to be analyzed.
  - tendency: string. The desired central tendency.
  Can be 'mean', 'median', or 'mode'. 'mean' gives
  uniform average by default and weighted mean with
  weight vector 'w' otherwise.
  - **arg:
    - w: array_like.

  Returns:
  - out: float. The desired central tendency.
  '''

  # making sample column vectors
  Data = preprocess(data)

  df = pd.DataFrame(data)

  if tendency == 'mean':

    if arg.get('w'):
      # making w a column vector
      W = weight_preprocess(arg.get('w'))
      out = np.dot(Data, W)[0 ,0]
      return out

    # uniform weights
    W = np.ones((Data.shape[1], 1))/Data.shape[1]
    out = np.dot(Data, W)[0, 0]
    return out

  if tendency == 'median':
    out = df.median()[0]
    return out

  if tendency == 'mode':
    out = df.mode()[0]
    return out
```

Code 1: Central tendency function for **C 1.2**

# 3 Extra Practice

## E 1.1

> Let $S = \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix} \right\}$ and vector $v = \begin{bmatrix} 9 \\ 16 \\ 29 \end{bmatrix}$. $v \in \text{span}(S)$ is true or false ? Why?

Assume we have the equations:

$$\begin{bmatrix} 9 \\ 16 \\ 29 \end{bmatrix} = a_1 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + a_2 \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} + a_3 \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix}$$

Then we get:

$$\begin{cases} 9 = a_1 + 2a_2 + a_3 \\ 16 = 2a_1 + 4a_2 + a_3 \\ 29 = 3a_1 + 6a_2 + 4a_3 \end{cases}$$

So,

$$v = (7 - a_2) \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + a_2 \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} + 2 \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix}$$

For all $a_2 \in \mathbb{R}$, we can conclude by definition that $v \in span(s)$

## E 1.2

> Let $S = \{v_1, \ldots, v_n\}$ be a set of vector in $V$. Prove that span$(S)$ is the smallest vector space of $S$, i.e
> - span$(S)$ is a subspace of $V$ and $S$ is a subset of span$(S)$
> - Any subspace $W$ of $V$, which contains $S$ , will be span$(S)$

We have $S = \{v_1, \ldots, v_n\} \in V$ and $span(S) = \{v = a_1 v_1 + \cdots + a_n v_n, a_1, \ldots, a_n \in \mathbb{R}\}$.

$span(S)$ is obviously the subspace of $V$ as $V$ closes under vector addition and scalar multiplication, and it is also obvious that $S$ is a subset of $span(S)$ because $span(S)$ is the set containing all the possible linear combinations of $S$.

Now, suppose that $W$ contains $S$, and since $W$ itself is a vector space, then it also contains every linear combinations of $S$ ($W$ is closed under vector addition and scalar multiplication), which happens to be $span(S)$. We can then conclude that $W$ has to be not smaller than $span(S)$, thus making $span(S)$ the smallest subspace that contains $S$.

### E 1.3

> Prove that $\{\emptyset\}$ is a independent linear set.

A vector space $V$ is said to be linearly dependent if

$$\sum_{i=0}^{n} a_i v_i = 0$$

where in $a_1, a_2, \ldots, a_i$ are scalars, not all zero, and $v_1, v_2, \ldots, v_i \in S$. Now we can conclude that $\varnothing$ is linearly independent, because simply it does not have any collection of $a$ that contains at least 1 non-zero scalars.

### E1.4

> Prove that $\{\mathbf{0}\}$ is a dependent set, then every set containing $\mathbf{0}$ too.

In the case of $\{0\}$, $\exists a \neq 0$ such that $a\vec{0} = \vec{0}$. Therefore it is linearly dependent by definition. The same applies to any vector space that contains $\vec{0}$, as in the collection of scalars $a_1, a_2, \ldots, a_i$ of said vector space $S = \{v_1, v_2, \ldots, v_n\}$, $\exists a_1 \neq 0, a_2, ..., a_n = 0$, which makes $a_1 0 + a_2 v_2 + \ldots + a_n v_n = 0$ (we can safely assume that $v_1 = 0$ as it does not break the generality of the statement).

## E 1.5

> Suppose that $V$ is a vector space with basis $B = (v_1, \ldots, v_n)$. Prove that $\forall v = \sum_{i=1}^{n} a_i v_i$ with $(a_1, \ldots, a_n)$ unique.

Assume for contradiction that there exists a $(b_1, \ldots, b_n)$ $\exists i \in \{1, \ldots, n\}$ s.t. $b_i \neq a_i$ $v = \sum_{i=1}^{n} b_i v_i$ besides $(a_1, \ldots, a_n)$ , $v = \sum_{i=1}^{n} a_i v_i$. Then we will have the following:

$$\sum_{i=1}^{n} b_i v_i = v = \sum_{i=1}^{n} a_i v_i$$

Equals to:

$$\sum_{i=1}^{n} (a_i - b_i) v_i = 0$$

Thus

$$a_i - b_i = 0 \Leftrightarrow a_i = b_1$$

contradicting the $b_i \neq a_i$ assumption. We then can conclude that the collection $(a_1, \ldots, a_n)$ , $v = \sum_{i=1}^{n} a_i v_i$ is unique.

## E 1.6

> Prove that probability simplex is also a vector space under appropriate definitions of vector addition and scalar multiplication.

### Hint

- Prove that there exists a bijection from $\mathbb{R}^{\ltimes}$ to $(0, \infty)^n$
- Contribute a appropriate addition and scalar multiplication on $(0, \infty)^n$ then the bijection is isormorphism between 2 vector spaces.
- Quotient of vector space is vector space (inherit property). Find $W$ is subspace of $(0, \infty)^n$ such that $f : (0, \infty)^n / W \mapsto \Delta$.

The proof on this link: https://golem.ph.utexas.edu/category/2016/06/how_the_simplex_is_a_vector_sp.html#c050689