VMLS - Study Guide Chapter 5 Name: Ben Chen

Always include this title page with your PDF. Include your name above.

- Submit your work in Gradescope as a PDF - you will identify where your "questions are."
- Identify the question number as you submit. Since we grade "blind" if the questions are NOT identified, the work WILL NOT BE GRADED and a 0 will be recorded. Always leave enough time to identify the questions when submitting.
- One section per page (if a page or less) - We prefer to grade the main solution in a single page, extra work can be included on the following page.
- Long instructions may be removed to fit on a single page.
- **Do not start a new question in the middle of a page.**
- Solutions to book questions are provided for reference.
- You may NOT submit given solutions - this includes minor modifications - as your own.
- Solutions that do not show individual engagement with the solutions will be marked as no credit and can be considered a violation of honor code.
- If you use the given solutions you must reference or explain how you used them, in particular...

For full credit, EACH book exercise in the Study Guides must use one or more of the following methods and FOR EACH QUESTION. Identify the number the method by number to ensure full credit.

Method 1 - Provide original examples which demonstrate the ideas of the exercise in addition to your solution.

Method 2 - Include and discuss the specific topics needed from the chapter and how they relate to the question.

Method 3 - Include original Python code, of reasonable length (as screenshot r text) to show how the topic or concept was explored.

Method 4 - Expand the given solution in a significant way, with additional steps and comments. All steps are justified. This is a good method for a proof for which you are only given a basic outline.

Method 5 - Attempt the exercise without looking at the solution and then the solution is used to check work. Words are used to describe the results.

Method 6 - Provide an analysis of the strategies used to understand the exercise, describing in detail what was challenging, who helped you or what resources were used. The process of understanding is described.

Method #5

1. Pick a section of Chapter 5 to annotate. (20 pts)

Chapter 5

Linear independence

In this chapter we explore the concept of linear independence, which will play an important role in the sequel.

5.1 Linear dependence

A collection or list of n-vectors $a_1, \ldots, a_k$ (with $k \geq 1$) is called *linearly dependent* if

scale

$$\beta_1 a_1 + \cdots + \beta_k a_k = 0$$

key is beta

holds for some $\beta_1, \ldots, \beta_k$ that are not all zero. In other words, we can form the zero vector as a linear combination of the vectors, with coefficients that are not all zero. Linear dependence of a list of vectors does not depend on the ordering of the vectors in the list.

When a collection of vectors is linearly dependent, at least one of the vectors can be expressed as a linear combination of the other vectors: If $\beta_i \neq 0$ in the equation above (and by definition, this must be true for at least one i), we can move the term $\beta_i a_i$ to the other side of the equation and divide by β_i to get

$$a_i = (-\beta_1/\beta_i)a_1 + \cdots + (-\beta_{i-1}/\beta_i)a_{i-1} + (-\beta_{i+1}/\beta_i)a_{i+1} + \cdots + (-\beta_k/\beta_i)a_k.$$

The converse is also true: If any vector in a collection of vectors is a linear combination of the other vectors, then the collection of vectors is linearly dependent.

Following standard mathematical language usage, we will say "The vectors $a_1, \ldots, a_k$ are linearly dependent" to mean "The list of vectors $a_1, \ldots, a_k$ is linearly dependent". But it must be remembered that linear dependence is an attribute of a *collection* of vectors, and not individual vectors.

This means that you can abstract away redundant info. Information like "Date of Birth" is enough to determine "Age", and "Year of birth". Age and Year of birth are both linearly dependent on DoB because that is information that can be obtained from evaluating the DoB. You can take this a step further and compress data!

Linear independent is less of an extension of the existing vector, but a way to cover completely different part of the grid (if in 2D space)

Linearly independent vectors. A collection of n-vectors $a_1, \ldots, a_k$ (with $k \geq 1$) is called *linearly independent* if it is not linearly dependent, which means that

$$\beta_1 a_1 + \cdots + \beta_k a_k = 0 \tag{5.1}$$

Method #5

2. (20 pts) Solve and explain the solution to 5.1 here in your own words. (Since you are given a solution, you will be graded on your ability to explain).

5.1 *Linear independence of stacked vectors.* Consider the stacked vectors

$$c_1 = \begin{bmatrix} a_1 \\ b_1 \end{bmatrix}, \quad \ldots, \quad c_k = \begin{bmatrix} a_k \\ b_k \end{bmatrix},$$

where $a_1, \ldots, a_k$ are n-vectors and $b_1, \ldots, b_k$ are m-vectors.

(a) Suppose $a_1, \ldots, a_k$ are linearly independent. (We make no assumptions about the vectors $b_1, \ldots, b_k$.) Can we conclude that the stacked vectors $c_1, \ldots, c_k$ are linearly independent?

(b) Now suppose that $a_1, \ldots, a_k$ are linearly dependent. (Again, with no assumptions about $b_1, \ldots, b_k$.) Can we conclude that the stacked vectors $c_1, \ldots, c_k$ are linearly dependent?

a) assume $x_1 a_1 + \dots + x_k a_k = 0$

this means $x_1 = x_2 = \dots = x_k = 0$

Stacked vectors look like $x_1 c_1 + \dots + x_k c_k = 0$

so $x_1 \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} + \dots + x_k \begin{bmatrix} a_k \\ b_k \end{bmatrix}$ has to $= \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$x_1 a_1 + \dots + x_k a_k = 0$ where they are $\mathbb{R}^n$

$x_1 b_1 + \dots + x_k b_k = 0$ & in $\mathbb{R}^n$

$0 \cdot b_1 + 0 \cdot b_2 + \dots + 0 \cdot b_k = 0$

Only possible solution to stacked vector is where $x_1 = x_2 = \dots x_k = 0$ and $\therefore$ linearly independent

b) if $a_1, \dots a_k$ are linearly dependent & scalars $x_1, x_2, x_3, \dots, x_k$ not all zeros such that $x_1 a_1 + \dots x_k a_k = 0$

so stacked vector of $x_1 c_1 + \dots x_k c_k = 0$

① $x_1 a_1 + \dots + x_k a_k = 0$ (proven in part a))

② $x_1 b_1 + \dots + x_k b_k = 0$ ← not sure but first equation confirms that $x_1 \dots x_k$ are not all zero which means we have at least 1 solution $\therefore$ $c_1, \dots, c_k$ is linearly independent

Method #5

3. (20 pts) Solve and Explain the solution to 5.2 here in your own words. (Since you are given a solution, you will be graded on your ability to explain).

5.2 *A surprising discovery.* An intern at a quantitative hedge fund examines the daily returns of around 400 stocks over one year (which has 250 trading days). She tells her supervisor that she has discovered that the returns of one of the stocks, Google (GOOG), can be expressed as a linear combination of the others, which include many stocks that are unrelated to Google (say, in a different type of business or sector).

Her supervisor then says: "It is overwhelmingly unlikely that a linear combination of the returns of unrelated companies can reproduce the daily return of GOOG. So you've made a mistake in your calculations."

Is the supervisor right? Did the intern make a mistake? Give a very brief explanation.

I love this question. The supervisor is wrong. When the intern says that GOOG can be expressed as a linear combination of other stocks, inclusive of ones that are unrelated to Google, she's right! This doesn't mean that other stocks can be used to predict Google's returns, but it means that Google's returns can be represented using the other stock returns due to having more stocks than days. You have more questions than unique sets of answers, which means that some of the questions must be expressible as a combination of others - not necessarily because there's meaningful connection between these data points, but simply because there are not enough unique trading days to make every stock completely independent of each other.

This is known as the pigeonhole principle of linear algebra. When you have more stocks than trading days, linear dependence is guaranteed so the intern's findings were spot on.

4. (ungraded) Solve and Explain the solution to 5.3 here in your own words. (Since you are given a solution, you will be graded on your ability to explain).

5. (ungraded 0 pts) Solve and Explain the solution to 5.4 here in your own words. (Since you are given a solution, you will be graded on your ability to explain).

Method #5

6. (20pts) Solve and Explain the solution to 5.5 here in your own words. (Since you are given a solution, you will be graded on your ability to explain).

5.5 *Orthogonalizing vectors.* Suppose that a and b are any n-vectors. Show that we can always find a scalar γ so that $(a - \gamma b) \perp b$, and that γ is unique if $b \neq 0$. (Give a formula for the scalar γ.) In other words, we can always subtract a multiple of a vector from another one, so that the result is orthogonal to the original vector. The orthogonalization step in the Gram–Schmidt algorithm is an application of this.

What scalar s.t. $(a - \gamma b) \perp b$?

vector

$$(a - \gamma b) \cdot b = 0$$

$$a \cdot b - \gamma (b \cdot b) = 0$$

$$\gamma = \frac{a \cdot b}{b \cdot b}$$ assume $b \neq 0$

property of transpose can turn this

$$(a - \gamma b)^T b \longrightarrow \gamma = \frac{a^T b}{b^T b}$$

$b^T b$ for clarity

This solution looks a lot like the one from the book, I did it without the "transpose" part of it, to make it simplier, but the transpose property allows for simple manipulation.

It makes sense that my solution looks like the books since there isn't much variance in the way you can solve this problem

Method #5

7. (20pts) Solve and Explain the solution to 5.6 here in your own words. (Since you are given a solution, you will be graded on your ability to explain).

5.6 *Gram–Schmidt algorithm.* Consider the list of n n-vectors

$$a_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad a_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \ldots, \quad a_n = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

(The vector a_i has its first i entries equal to one, and the remaining entries zero.) Describe what happens when you run the Gram–Schmidt algorithm on this list of vectors, *i.e.*, say what $q_1, \ldots, q_n$ are. Is $a_1, \ldots, a_n$ a basis?

$q_1 = a_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$

$q_2 = a_2 - \frac{a_2 q_1}{q_1 \cdot q_1} q_1$ $\qquad a_2 q_1 = 1 \quad q_1 q_1 = 1$

$q_2 = a_2 - q_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ \vdots \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \end{bmatrix}$

there is a pattern here with the "1" moving down the vector the more i it is and "1" appears in the "i"-th position

what about $q_1, \ldots q_n$? output ends up as an orthonormal standard basis vector of $\mathbb{R}^n$

$a_1, \ldots a_n$ is a basis because Gram-Schmidt transforms $a_1, \ldots a_n$ into a standard basis which means that original set was already a basis

they span the space and are linearly independent

I did Gram Schmidt by hand so I can see where "1" travels in the vector