EECS 245 Course Notes

Friday 11$^{\text{th}}$ July, 2025

1 Part 1: Introduction to Supervised Learning

1.1 1.1. Introduction to Machine Learning

What is machine learning? Supervised vs. unsupervised. Show the taxonomy.

Perhaps talk briefly about the history of machine learning.

When giving examples of supervised learning problems, talk a little bit about language models, like ChatGPT.

1.2 1.2. Loss Functions and the Constant Model

Learning Objectives

- Understand the idea of a model, loss function, and average loss.

The usual start from 398 Lecture 11.

1.3 1.3. Empirical Risk Minimization

Here, revisit the three-step modeling process from the previous section through the lens of absolute loss. Then talk about the minimizing outputs, i.e. $R_{\text{sq}}(h^*)$.

1.4 1.4. Simple Linear Regression

Upgrade to simple linear regression.

TODO need to show the loss surface in 3D and talk about partial derivatives.

Here, add that cool animation of partial derivatives where they "slice" into the 3D surface.

1.5 1.5. Least Squares

Mention, by name, the idea of least squares curve fitting, e.g. the parabola example from Wikipedia. (Mostly naming this “least squares” so that it appears in the sidebar.)

Motivate multiple linear regression and why vectors are actually needed. Relatively short section.

2 Part 2: Vectors and Matrices

2.1 2.1. Vectors

2.1.1 Introduction

Vectors are the central object of study in linear algebra, and appear in many different contexts. Thanks to the last note, you have *some* understanding as to why they might be relevant in our journey through machine learning – they are used to store data, and in solving systems of linear equations.

We'll start with the most basic and practically relevant definition of a vector. We'll introduce more abstract definitions later, as they become relevant.

Definition: Vector
A **vector** is an ordered list of numbers.

In this class, we'll typically use lowercase letters to denote vectors, drawn with arrows above the letters. For example:

$$\vec{v} = \begin{bmatrix} 4 \\ -3 \\ 15 \end{bmatrix} \tag{1}$$

By *ordered list*, we mean that the order of the numbers in the vector matters.

- For example, the vector $\vec{v} = \begin{bmatrix} 4 \\ -3 \\ 15 \end{bmatrix}$ is not the same as the vector $\vec{w} = \begin{bmatrix} 15 \\ -3 \\ 4 \end{bmatrix}$, despite the fact that they have the same components.
- $\vec{v}$ is also different from the vector $\vec{u} = \begin{bmatrix} 4 \\ -3 \\ 15 \\ 1 \end{bmatrix}$, despite the fact that their first three components are the same.

In general, we're mostly concerned with vectors in R^n, which is the **set** of all vectors with n **components** or **elements**, each of which is a real number. It's possible to consider vectors with complex components (the set of all vectors with complex components is denoted C^n), but we'll stick to real vectors for now.

The vector $\vec{v}$ defined in the box above is in R^3, which we can express as $\vec{v} \in R^3$.

A general vector in R^n can be expressed by:

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \tag{2}$$

Before we move forward, let's clarify the notation we're using for subscripts. In the definition of $\vec{v}$ above, the components of the vector are denoted $v_1, v_2, \ldots, v_n$. Each of these individual components is a **single** real number, not a vector. But in the near future, we may want to consider multiple vectors at once, and may use subscripts to refer to them as well:

$$\vec{v}_1 = \begin{bmatrix} 4 \\ -3 \\ 15 \end{bmatrix} \quad \text{and} \quad \vec{v}_2 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \tag{3}$$

The meaning of the subscript depends on the context, so just be careful!

That's all great. **But, what do these numbers mean?** Vectors encode remarkable amounts of information and beauty, and shouldn't be thought of as "just" a list of numbers.

It turns out that we have a solution for running code directly in this note, without you needing to open Vocareum. Check it out below:

2.1.2 Norm (i.e. Length or Magnitude)

In the context of physics, vectors are often described as creatures with "a magnitude and a direction". While this is not a physics class – this is EECS 245, after all! – this interpretation will be useful for us, too.

To illustrate what we mean, let's consider a concrete vector in R^2, since it is easy to visualize vectors in 2 dimensions on a computer screen. Suppose $\vec{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$. Then, **geometrically**, we can visualize $\vec{v}$ as an arrow pointing from the origin $(0, 0)$ to the point $(3, 4)$ in the 2D Cartesian plane.

The vector $\vec{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ moves 3 units to the right and 4 units up, which we know by reading the components of the vector.

It's worth noting that $\vec{v}$ isn't "fixed" to start at the origin – vectors don't have positions. All three vectors in the figure below are the same vector, $\vec{v}$.

TODO add a visualization of the vector $\vec{v}$ in 3 different positions.

What's not immediately obvious is how "long" the vector is. To compute this vector's length, we should remember the Pythagorean theorem, which states that if we have a right triangle with legs of length a and b, then the length of the hypotenuse is $\sqrt{a^2 + b^2}$. In the example above, we have a right triangle with legs of length 3 and 4, so the length of the hypotenuse is $\sqrt{3^2 + 4^2} = \sqrt{25} = 5$, so the vector above has length 5.

Great, so we know that $\vec{v}$ travels 5 units. But in what direction does $\vec{v}$ travel? We'll address this point in just a moment. But first, let's generalize the calculation we just performed.

Definition: Vector Norm
The **norm** of a vector $\vec{v} \in R^n$ is defined as follows:

$$\|\vec{v}\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2} \tag{4}$$

The norm of a vector is also called its **length** or **magnitude**. This particular formula for the norm is also called the **Euclidean norm** or L_2 norm, and is the most common and "default" norm used in linear algebra. In coming lectures, we'll see other norms, which describe different ways of measuring the "length" of a vector.

What may not be immediately obvious is *why* the Pythagorean theorem seems to extend to higher dimensions – the 2D case seems reasonable, but why is the length of the vector $\vec{u} = \begin{bmatrix} 2 \\ -3 \\ 6 \end{bmatrix}$ in R^3 equal to $\sqrt{2^2 + (-3)^2 + 6^2}$? The cop-out answer is that this is a definition, and I *could* define the length of a vector to be whatever I'd like it to be, but we should verify that this definition makes intuitive sense and has reasonable properties. Let's consider that same vector $\vec{u}$ for a moment.

TODO: Draw a picture of the vector $\vec{u}$ in 3D space.

There are actually two right angle triangles in the picture above:

- One triangle has legs of length 2 and 3, with a hypotenuse of h.
- Another triangle has legs of length h and 6, with a hypotenuse of $\|\vec{u}\|$.

To find $\|\vec{u}\|$, we can use the Pythagorean theorem twice:

$$h^2 = 2^2 + (-3)^2 = 4 + 9 = 13 \implies h = \sqrt{13} \tag{5}$$

Then, we can use the Pythagorean theorem again to find $\|\vec{u}\|$:

$$\|\vec{u}\| = \sqrt{h^2 + 6^2} = \sqrt{\sqrt{2^2 + (-3)}^2 + 6^2} = \sqrt{2^2 + (-3)^2 + 6^2} = \sqrt{49} = 7 \tag{6}$$

So, to find $\|\vec{u}\|$, we used the Pythagorean theorem twice, and ended up computing the square root of the sum of the squares of the components of the vector, which is what the definition above states. This argument naturally extends to higher dimensions.

> **Warning**
> Do not confuse the length of a vector with the number of components in a vector!
> If $\vec{v} \in R^n$, then the length of $\vec{v}$ is $\|\vec{v}\|$, while the number of components in $\vec{v}$ is n.

Vector norms satisfy several interesting properties, which we will introduce later in this note once we have a bit more context.

2.1.3 Direction

Let's return to the vector $\vec{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$. We know that $\|\vec{v}\| = 5$. But in what direction does $\vec{v}$ travel?

TODO show $\vec{v}$ once again.

Directions are *relative*, and it's standard to describe directions in terms of the **angle** they make with the positive x-axis. Suppose the angle between $\vec{v}$ and the positive x-axis is θ. Then, we can use trigonometry to find θ in a few equivalent ways:

- $5\cos\theta = 3 \implies \cos\theta = \frac{3}{5} \implies \theta = \arccos\left(\frac{3}{5}\right) \approx 0.9273 \text{ radians} = 53.13°$.
- $5\sin\theta = 4 \implies \sin\theta = \frac{4}{5} \implies \theta = \arcsin\left(\frac{4}{5}\right) \approx 0.9273 \text{ radians} = 53.13°$.
- $\tan\theta = \frac{4}{3} \implies \theta = \arctan\left(\frac{4}{3}\right) \approx 0.9273 \text{ radians} = 53.13°$.

We can now fully describe $\vec{v}$ in two ways:

- It is the vector $\vec{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$.
- It is the vector in R^2 with length 5 and angle 53.13° with the positive x-axis.

Both of these descriptions are equivalent and uniquely describe the same $\vec{v}$.

We'll have more to say about angles shortly, when we discuss the angle **between** two vectors, but we figured a review of the basics couldn't hurt.

2.1.4 Vector Addition and Scalar Multiplication

Out of the box, vectors support two basic operations: addition and scalar multiplication. These relatively simple operations will allow us to study sophisticated behavior.

Definition: Vector Addition
Suppose $\vec{u}$ and $\vec{v}$ are both vectors with the same number of components, i.e. $\vec{u}, \vec{v} \in R^n$.

Then, the **sum** of $\vec{u}$ and $\vec{v}$ is defined as follows:

$$\vec{u} + \vec{v} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_n \end{bmatrix} \tag{7}$$

This tells us that vector addition is performed **element-wise**. This is a term that you'll encounter quite a bit in the context of writing `numpy` code, as you'll see in lab.

For example, if $\vec{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$, then $\vec{u} + \vec{v} = \begin{bmatrix} 5 \\ 7 \\ 9 \end{bmatrix}$.

If $\vec{w} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, then $\vec{u} + \vec{w}$ is undefined, since $\vec{u}$ and $\vec{w}$ have different numbers of components.

Definition: Scalar Multiplication
Suppose $\vec{v} \in R^n$. The **scalar multiple** of $\vec{v}$ by a scalar $c \in R$ is defined as follows:

$$c\vec{v} = c \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} cv_1 \\ cv_2 \\ \vdots \\ cv_n \end{bmatrix} \tag{8}$$

For example, if $\vec{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $c = 2$, then $2\vec{v} = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}$. Note that we've deliberately defined this operation as **scalar** multiplication, not just "multiplication" in general, as there's more nuance to the definition of multiplication in linear algebra.

The norm of a vector satisfies the following properties, which you should verify for yourself:

- $\|\vec{v}\| \geq 0$.

- $\|\vec{v}\| = 0$ if and only if $\vec{v} = \vec{0} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$.
- $\|c\vec{v}\| = |c|\|\vec{v}\|$ for all $c \in R$.
- $\|\vec{u} + \vec{v}\| \leq \|\vec{u}\| + \|\vec{v}\|$ (called the **triangle inequality**; the proof of this is non-trivial, and we'll see it later in a homework assignment).

2.1.5 Visualizing Addition and Scalar Multiplication

2.2 2.2. Dot Product and Orthogonality

2.2.1 Introduction

In the last note, we introduced two key operations that work on vectors:

- vector addition.
- scalar multiplication.

We also introduced the notion of a vector norm, which is a measure of the "length" of a vector.

Here, we'll introduce another key.

2.2.2 The Dot Product

Suppose $\vec{u}, \vec{v} \in R^n$. The **dot product** of $\vec{u}$ and $\vec{v}$ is defined as follows:

2.3 2.3. Linear Independence

What is a linear combination? Give examples with one vector, multiple vectors. What do 2 vectors in R3 span? What do 5 vectors in R2 span? etc.

Motivate by saying vectors contain information about individuals, or products, and we want to create formulas that combine them in some linear fashion.

Talk about linear independence, and focus on counting the number of linearly independent vectors in a set.

Good site to borrow problems from: https://mmids-textbook.github.io/chap02_ls/exercises/roch-mmids-ls-exercises.html

2.4 2.4. Projection, Part 1

Talk about projecting onto the span of a single vector.

Talk about span of multiple vectors.

2.5 2.5. Matrices

Shapes of matrices – draw a long, wide, square, etc.

Matrix multiplication, pull from 398 note. Talk about matrix-vector multiplication as being a linear combination of the columns of the matrix, with the weights being the entries of the vector.

Talk about the rank of a matrix being the number of linearly independent columns. CR factorization and how that relates to rank.

Define the matrix transpose, and define another way for defining the dot product.

2.6 2.6. Linear Transformations

More on matrix-vector multiplication as being an operator that sends $\vec{x}$ to $A\vec{x}$, and what various A do to $\vec{x}$.

Other topics:

- Inverse
- Determinant

2.7 2.7. Vector Spaces

Four fundamental subspaces of a matrix: column space, null space, row space, left null space.

These are really just the column space of null space of both A and A^T.

Rank-nullity theorem, if not in the previous note.

Subspaces, and how the span of a set of vectors in R^n is a subspace of R^n.

What is a basis; homework problem: finding an orthonormal basis for the column space of a matrix.

Later: inner product space.

2.8 2.8. Projection, Part 2

Projecting onto the column space of a matrix, i.e. the span of the columns of a matrix.

The classic geometric picture.

3 Part 3: Regression using Linear Algebra

3.1 3.1. From Scalars to Vectors

Recap of linear regression and loss functions.

Define hypothesis vector, parameter vector, error vector, design matrix, etc.

3.2 3.2. Incorporating Multiple Features

Incorporating multiple features.

3.3 3.3. Overfitting

Overview of model complexity, overfitting and underfitting, the bias-variance tradeoff (at least in the classical sense).

4 Part 4: Gradients

4.1 4.1. The Gradient Vector

Recap of partial derivatives. Give a visual sense of what the gradient vector is.

Emphasize the importance of understanding the domain and range of a function, i.e.

Rn̂ - R, R - R, etc.

Give examples of gradients for xTx, xT Ax, etc.

4.2 4.2. Least Squares, Revisited

Now that we know how to take the gradient of a function Rn̂ ->R, let's minimize

$$R_{\text{sq}}(\vec{w}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \vec{w}^T\vec{x}_i)^2 \tag{9}$$

by taking the gradient and setting it to 0.

This could have been part of the previous note, but it's important enough to have its own note.

4.3 4.3. Convexity

Talk about convexity in the single-variable sense, local vs. global minima, linear approximations, Taylor Series, etc.

4.4 4.4. Gradient Descent

Gradient descent, first in the single-variable sense, then in the multi-variable sense.

5 Part 5: Eigenvalues and Eigenvectors

5.1 5.1. Eigenvalues and Eigenvectors

Include PageRank as a motivating example. But make sure to talk about linear transformations and the geometric view.

Somewhere in here: case study of rotation matrices.

5.1.1 Eigenvalues and Eigenvectors

Learning Objectives
TODO

Websites and Networks As we did at the start of the course, we'll start with an example that is rooted in application, and introduce the necessary mathematical machinery to study it. At the start of the course, we studied the ever-present problem of building models that make meaningful predictions about the future – and rest assured, we will return to that problem shortly. For now, though, we'll start by studying the problem that led to the creation of Google, which you may have once heard of by the phrase "the billion dollar eigenvalue."

The Punchline On its own, this definition is a bit hard to parse. But, it really says something quite simple and profound:

Important
If A is a matrix, and $\vec{v}$ is an eigenvector of A, then $A\vec{v}$ is a **vector that points in the same direction** as $\vec{v}$!

Let's put that another way. The function $f(\vec{v}) = A\vec{v}$ is a **linear transformation** that maps vectors in R^n to vectors in R^n. Eigenvectors of A are vectors whose directions are **unchanged** by the linear transformation f. The corresponding eigenvalues are the scalars by which the eigenvectors are scaled.

5.2 5.2. Eigendecomposition

Recap of eigenvalues and eigenvectors.

Writing $A = V\Lambda V^{-1}$ is a way to decompose a matrix into a diagonal matrix of eigenvalues and a matrix of eigenvectors.

Talk about the rank of a matrix and how that relates to its eigenvalues.

Positive semidefinite matrices, and how they relate to the eigenvalues of a matrix. Re-introduce least squares from this perspective.

5.3 5.3. Singular Value Decomposition

Remember, only square matrices have eigenvalues and eigenvectors.

For non-square matrices, we can use the singular value decomposition, which has similar properties and will enable for some powerful data analysis.

5.4 5.4. Principal Components Analysis

Recap SVD, talk about how it relates to PCA.

Show the taxonomy of machine learning, and how PCA is a dimensionality reduction technique, which is a form of unsupervised learning.

6 Part 6: Probability

6.1 6.1. Introduction to Random Variables

Don't need to do the first-principles definitions, can define a random variable intuitively.

6.2 6.2. Continuous Distributions

Introduce the notion of a pdf and areas as being probabilities. Uniform and normal Gaussian RVs. Emphasis on intuition. Covariance matrices.

6.3 6.3. Independence

Bayes' rule in depth. The definition of conditional independence, and how it differs from that of independence. Law of total probability.

6.4 6.4. Maximum Likelihood Estimation

Start with estimating the bias of a coin (i.e. estimating p in n samples of Bern(p)), and work up to more sophisticated examples, all in the context of discrete RVs. Relates back to calculus and optimization (which is why we've chosen MLE over other probability ideas). Maximizing likelihood vs. log-likelihood.

6.5 6.5. Least Squares, Revisited, Again

Motivate squared loss through the lens of MLE.

7 Part -1: Math Review

This course will rely on your understanding of some ideas from calculus and high school algebra. This section of the notes serves to review these ideas, and to introduce some notation that we'll use throughout the course.

Regardless of your background, make sure to skim the content of this section, and attempt the exercises if these ideas are new to you. Core course content starts in Part 1: Supervised Learning.

7.1 Summation Notation and the Mean

Sums and averages play an important role in machine learning. In Part 1 of the course, we'll learn to take the average of an important measurement (called a "loss function") for every value in our dataset.

Here, we'll review the most relevant properties of summation notation, and use the arithmetic mean as a case study of sorts. In the next note, on Calculus, we'll revisit summation notation in the context of derivatives.

7.1.1 Introduction

Definition: Summation Notation

The $\sum$ symbol, read "sigma", is used to indicate a sum of a sequence. In general, if a and b are integers, and f is a function, then:

$$\sum_{i=a}^{b} f(i) = f(a) + f(a+1) + f(a+2) + \quad + f(b-1) + f(b) \tag{10}$$

Above, i is the *index of summation*.

For example, $\sum_{i=1}^{6} i^2$ represents the sum of the squares of all integers from 1 to 6:

$$\sum_{i=1}^{6} i^2 = 1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2 = 91 \tag{11}$$

Notice that both the starting and ending indices (1 and 6, respectively) are included in the sum.

Often, we'll take the sum of the first n terms of a sequence. The sum of the squares of the first n positive integers is:

$$\sum_{i=1}^{n} i^2 = 1^2 + 2^2 + ... + (n-1)^2 + n^2 \tag{12}$$

The example above illustrates **why** we use summation notation – it allows us to express the sum of a sequence in a compact way, as it prevents us from having to write out each term individually. (In this most recent example, we don't know the actual value of n, we wrote out just the first few and last few terms to unambiguously define the pattern, and then used ... to indicate that the pattern continues.)

Note that the index of summation can be any variable name (i is just a typical choice). That is, $\sum_{j=1}^{n} j^2$, $\sum_{i=1}^{n} i^2$, and $\sum_{\text{zebra}=1}^{n} \text{zebra}^2$ all represent the same sum.

Summation notation can be thought of in terms of a `for`-loop. In Python, to compute the sum $\sum_{i=a}^{b} f(i)$, we could write:

```
total = 0
for i in range(a, b + 1):
 total = total + f(i)
```

As we mentioned above, the ending index is *inclusive* in summation notation. This is in contrast to Python, where the ending index is *exclusive*, which is why we provided `b + 1` as the second argument to the `range` function instead of `b`.

7.1.2 Properties

Constants The sum of the heights

> **Warning**
> Summation notation **does not distribute over division**. That is,
>
> $$\sum_{i=1}^{n} \frac{x_i}{y_i} \neq \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} y_i} \tag{13}$$

- Pulling a constant
- Separating sums
- Re-indexing
- Variable scope
- Infinite series
- Telescoping sums as an example

Something about:

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n \tag{14}$$

Provide lots of exercises.

7.1.3 Mean and Standard Deviation

TODO put this after the properties come

Suppose we have a sequence of n numbers, x_1, x_2, ..., x_n. Perhaps these represent the heights of all n students in our class. The **mean**, or **average**, of all n values is given the symbol $\bar{x}$ (pronounced "x-bar") and is defined as follows:

$$\bar{x} = \frac{x_1 + x_2 + ... + x_{n-1} + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{15}$$

Above, we've used our new notation for the sum of a sequence.

You've likely seen this definition before. But, an often-forgotten property of the mean is that the **sum of the deviations from the mean is zero**. By that, I mean (no pun intended) that if you:

1. compute the mean of a sequence of numbers,
2. compute the *signed* difference between each number and the mean, and then
3. sum all of those differences, the result will be zero.

Let's first see this in action, then show why it is true in general. Suppose there are only 4 students in the class, with heights 72, 63, 68, and 65 inches. The mean of these heights is:

$$\bar{x} = \frac{72 + 63 + 68 + 65}{4} = 67 \tag{16}$$

The deviations from the mean are:

$$\begin{aligned} 72 - 67 &= 5 \\ 63 - 67 &= -4 \\ 68 - 67 &= 1 \\ 65 - 67 &= -2 \end{aligned}$$

The sum of the four deviations, then, is:

$$-5 + (-4) + 1 + (-2) = 0 \tag{17}$$

So, the mean deviation from the mean is zero in this example.

This is also true in general. Precisely, I'm claiming that if $x_1, x_2, ..., x_{n-1}, x_n$ are any n numbers, and $\bar{x}$ is their mean, then $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$.

Let's prove it. I'll start with the left-hand side of the equation:

$$\sum_{i=1}^{n}(x_i - \bar{x}) \tag{18}$$

7.2 Derivatives

In Part 1, we'll rely on a few key ideas from a first course in calculus (e.g. Math 115). Before proceeding, make sure you've read the previous note on Summation Notation, as we'll use some of the notation from there.

TODO put example of single variable function with local and global minima, talk about how we find them.

Derivatives of sums.

7.3 Logarithms

It may seem strange to have an entire page dedicated to logarithms, but the humble function $f(x) = \log(x)$ plays a central role in the study of machine learning. Here, we'll review some of its properties, and look at how the previous two notes – on summation notation and calculus – play a role.

Definition

Talk about the base

Properties

Summation

7.4 Sets

In linear algebra, we often study properties of **sets** of vectors, so it'll help to be familiar with the basic ideas of set theory. If you've taken EECS 203, this should all be very familiar.

https://notes.imt-decal.org/sets/sets-and-set-operations.html

At the end, list out the sets of numbers students should be familiar with: https://notes.imt-decal.org/sets/sets-of-numbers.html (no need to discuss cardinality)