NMEP

as a Linear Combination of
Math 54
(with lots of coefficients set to 0)

1 Vectors

1.1 Lists of numbers

Some vectors (but only **some**) are lists of numbers:

$$\begin{bmatrix} 1 & 2 & 3 & \cdots & n \end{bmatrix}$$

For now, we will focus on this type of vectors, but remember that it is a very particular special case.

1.2 Typical notation

We use $\overrightarrow{x}$ to unambiguously denote x as a vector variable. However, we will also use bold $\mathbf{x}$ (and simply x in hand-writing), whenever the nature of x is obvious from the context.

We use $\overrightarrow{x} \in \mathbb{R}^n$ or, more generally, $\overrightarrow{x} \in \mathbb{R}^{n\times 1}$ to denote that vector $\overrightarrow{x}$ has n elements. Ideally get used to thinking of n as the **height** of the vector and not its length. Also note that $\mathbb{R}$ in this notation means that this vector is particularly a list of real numbers. For non-numerical vectors we sometimes have no choice but to be verbose and type out that "$\overrightarrow{x}$ is a vector of n elements'.

NOTE: You will sometimes hear engineers abuse Computer Science slang and say "length" of a vector to describe the number of its elements. They are all going to math jail.

For numerical ($\mathbb{R}^n$) vectors, *norm* is defined as the euclidean distance of their endpoint from the origin. The notation for norm of $\overrightarrow{x}$ is $\|\overrightarrow{x}\|$. Sometimes, especially in hand-writing, we use just $|\overrightarrow{x}|$ (this notation conflicts with the absolute value and so is discouraged).

$$\|\overrightarrow{x}\| = \| \begin{bmatrix} 1 & 2 & 3 & \cdots & n \end{bmatrix} \| \coloneqq \sqrt{1^2 + 2^2 + 3^2 + \ldots + n^2}$$

NOTE: People who prefer geometry to abstract algebra sometimes say *length* to refer to norm. They have connections in the math department and so do not go to math jail.
Engineers are left confused and slighted.

NOTE: Norm is also defined for non-numerical vectors. **Euclidean distance is just the most common special case of the definition of norm.** In fact, other norms can be used for $\mathbb{R}^n$ too.
See https://en.wikipedia.org/wiki/Norm_(mathematics)

We write vectors horizontally (slyly pretending that they are *row vectors*) only if we just want to list the vector's contents. In all other cases (e.g. in equations) we write vectors vertically (as *column vectors*).

NOTE: We generally never use row vectors. The horizontal notation is just more convenient in writing. It still represents **column vectors**.

$$\overrightarrow{x} \in \mathbb{R}^{n\times 1} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

For alternative notation, check out https://en.wikipedia.org/wiki/Vector_notation

1.3 Numerical vector operations

$\vec{x}$ and $\vec{y}$ have an equal number of elements. Otherwise we cannot add or subtract them, or find their dot product.

$$\vec{x}, \vec{y} \in \mathbb{R}^n$$

Addition:

$$\vec{x} + \vec{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} 16 \\ 13 \\ 22 \\ \vdots \\ y_n \end{bmatrix} := \begin{bmatrix} 17 \\ 15 \\ 25 \\ \vdots \\ x_n + y_n \end{bmatrix}$$

Subtraction:

$$\vec{x} - \vec{y} = \begin{bmatrix} 20 \\ 13 \\ 3 \\ \vdots \\ x_n \end{bmatrix} - \begin{bmatrix} 16 \\ 2 \\ 22 \\ \vdots \\ y_n \end{bmatrix} := \begin{bmatrix} 4 \\ 11 \\ -19 \\ \vdots \\ x_n - y_n \end{bmatrix}$$

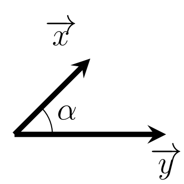
Dot product:
There are two equivalent (proof omitted) definitions/views of dot product.

The algebraic view:

$$\vec{x} \cdot \vec{y} = \begin{bmatrix} 4 \\ 8 \\ 16 \\ \vdots \\ x_n \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 5 \\ -7 \\ \vdots \\ y_n \end{bmatrix} := 4 + 40 + -112 + \ldots + x_n y_n = \sum_{i=1}^{n} x_n y_n$$

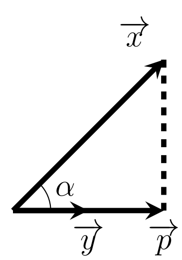
$$\begin{bmatrix} 5 \\ 9 \\ 17 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 6 \\ -8 \end{bmatrix} = 10 + 54 - 136 = -72$$

The geometric view:



$$\vec{x} \cdot \vec{y} := \|\vec{x}\|\|\vec{y}\| \cos(\alpha)$$

In particular, if $\|\vec{y}\| = 1$, the picture is even more convenient. This is related to vector projections which will be discussed later.



$$\boxed{\vec{x} \cdot \vec{y} = \|\vec{x}\| \cos(\alpha) = \|\vec{p}\|}$$

In three dimensions ($\vec{x}, \vec{y} \in \mathbb{R}^3$), there is also a notion of a *cross product*:

$$\vec{x} \times \vec{y} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \times \begin{bmatrix} a \\ b \\ c \end{bmatrix} := \begin{bmatrix} yc - zb \\ za - xc \\ xb - ya \end{bmatrix}$$

2 Linear Regression

This is a quick run through linear regression from both a linear algebra and a multivariable calculus perspective.

2.1 The Problem

$(\vec{x}_i, y_i)$ is a set of n samples (i.e. examples, also called *observations*). Each sample is a vector of m elements, called *features*. Let matrix X and vector $\vec{y}$ be vertical arrangements of these examples:

$$X = \begin{bmatrix} \vec{x}_1 \\ \vec{x}_2 \\ \vec{x}_3 \\ \vdots \\ \vec{x}_n \end{bmatrix} = \begin{bmatrix} \vec{x}_{11} & \vec{x}_{12} & \vec{x}_{13} & \cdots & \vec{x}_{1m} \\ \vec{x}_{21} & \vec{x}_{22} & \vec{x}_{23} & \cdots & \vec{x}_{2m} \\ \vec{x}_{31} & \vec{x}_{32} & \vec{x}_{33} & \cdots & \vec{x}_{3m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vec{x}_{n1} & \vec{x}_{n2} & \vec{x}_{n3} & \cdots & \vec{x}_{nm} \end{bmatrix}$$

$$\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}$$

Find the vector $\vec{w}$ such that $\|X \cdot \vec{w} - \vec{y}\|$ is minimized.

NOTE: $\|n\|$ may be the L1 norm, the L2 norm, or Ridge ($\|\vec{n}\| = \mathrm{L2}(\vec{n}) - \lambda\,\mathrm{L2}(\vec{w})$) or Lasso (($\|\vec{n}\| = \mathrm{L2}(\vec{n}) - \lambda\,\mathrm{L1}(\vec{w})$)) regression norms.

2.2 Linear Algebra Perspective

Consider an equation $X \cdot \vec{w_0} = \vec{y}$. This equation has a solution $\vec{w_0} = X^{-1} \cdot \vec{y}$ iff X is invertible.

Let X^{-1} be undefined. We know that we cannot solve the equation, but we will try to come up with a "best" possible solution.

Remember that the result of multiplying a matrix by a vector is a linear combination of the columns of the matrix by definition:

$$X \cdot \vec{w} = \begin{bmatrix} \vec{x_1} & \vec{x_2} & \vec{x_3} & \cdots & \vec{x_n} \end{bmatrix} \cdot \begin{bmatrix} (\vec{w})_1 \\ (\vec{w})_2 \\ (\vec{w})_3 \\ \vdots \\ (\vec{w})_n \end{bmatrix} = \vec{x_1}(\vec{w})_1 + \vec{x_2}(\vec{w})_2 + \vec{x_3}(\vec{w})_3 + \ldots + \vec{x_n}(\vec{w})_n$$

This means that no matter our choice of $\vec{w}$, it will be mapped by X into its column space:

$$X \cdot \overrightarrow{w} \in \text{Span}(\overrightarrow{x_1}, \overrightarrow{x_2}, \overrightarrow{x_3}, \ldots, \overrightarrow{x_n}) = \text{CS}(X)$$

This means that the best we can do if $\overrightarrow{y}$ does not lie in the column space of X is find $\hat{\mathbf{y}} \in \text{CS}(X)$ closest to $\overrightarrow{y}$.

Define a best approximation $\overrightarrow{w}$ of the solution $\overrightarrow{w_0}$ as the vector $\overrightarrow{w}$ such that $\|\hat{\mathbf{y}} - \overrightarrow{y}\| = \|X \cdot \overrightarrow{w} - \overrightarrow{y}\|$ is minimized.

As we know from high school geometry, the perpendicular is the shortest path between a point and a (hyper)plane (proof omitted). Thus, minimizing $\|\hat{\mathbf{y}} - \overrightarrow{y}\|$ is equivalent to finding $\hat{\mathbf{y}} \in \text{CS}(x)$ such that $(\hat{\mathbf{y}} - \overrightarrow{y}) \perp \text{CS}(X)$, which is trivially the projection of $\overrightarrow{y}$ onto the column space of X.

$$\hat{\mathbf{y}} = proj_{\text{CS}(X)} \overrightarrow{y}$$