# 1 Back to Basics: Linear Algebra

Let $X \in \mathbb{R}^{n \times m}$. We study a few important subspaces in the theory of linear maps. When we write $\subseteq$, it means "is a subspace of."

The **columnspace**, also called the range or span, of $X$ is $Range(X) := \{Xv : v \in \mathbb{R}^m\} \subseteq \mathbb{R}^n$. Consists of all vectors in the span (the set of all linear combinations) of the columns of $X$.
The **rowspace** is $Row(X) := \{X^\top v : v \in \mathbb{R}^n\} \subseteq \mathbb{R}^m$. Consists of all vectors in the span of the rows of $X$.
The **nullspace**, also called the kernel, of $X$ is $\mathcal{N}(x) := \{v \in \mathbb{R}^m : Xv = 0\} \subseteq \mathbb{R}^m$.
The **orthogonal complement** of a subspace $U$ in some vector space $V$ is a subspace, denoted $U^\perp$, such that $u \in U, v \in U^\perp \implies u \cdot v = 0$ and $U$ and $U^\perp$ together span $V$. (These facts imply that $\dim U + \dim U^\perp = \dim V$. It also implies that $U^{\perp\perp} = U$.) For example, in the three-dimensional Euclidean space $V = \mathbb{R}^3$, if $U$ is a plane through the origin, then $U^\perp$ is a line through the origin perpendicular to $U$.

For this problem we do not assume that $X$ has full rank.

(a) Show that the following facts are true.

(i) $Row(X) = Range(X^\top)$

(ii) $\mathcal{N}(X)^\perp = Row(X)$.

(iii) $\mathcal{N}(X^\top X) = \mathcal{N}(X)$ *Hint: if $v \in \mathcal{N}(X^\top X)$, then $v^\top X^\top X v = 0$.*

(b) We now prove an important result of linear algebra, the rank-nullity theorem. Let $Rank(X) = \dim Range(X) = \dim Row(X)$ and $Nullity(X) = \dim \mathcal{N}(X)$. (The fact that $\dim Range(X) = \dim Row(X)$—that is, the dimension spanned by the rows equals the dimension spanned by the columns—is itself a pretty important result, which you should always remember when you hear the word "rank.") The rank-nullity theorem says that for any $X \in \mathbb{R}^{nxm}$,

$$Rank(X) + Nullity(X) = m.$$

Use the above results to prove this theorem. *Hint: Use the orthogonal complement of the nullspace to connect the rank to the nullity.*

 

Gilbert Strang has proposed that a collection of four facts be called "fundamental theorem of linear algebra." Two of these facts are the rank-nullity theorem, part (b), and the fact that the rowspace is the orthogonal complement of the nullspace, part (a)(ii). The other two facts are related to the singular value decomposition, which we'll learn late in the semester.

 

## 2 Eigenvalues

(a) Let $\mathbf{A}$ be an invertible matrix. Show that if $\mathbf{v}$ is an eigenvector of $\mathbf{A}$ with eigenvalue $\lambda$, then it is also an eigenvector of $\mathbf{A}^{-1}$ with eigenvalue $\lambda^{-1}$.

(b) A symmetric matrix $\mathbf{A}$ is said to be positive semidefinite (PSD) ($\mathbf{A} \geq 0$) if $\forall \mathbf{v} \neq 0, \mathbf{v}^\top \mathbf{A} \mathbf{v} \geq 0$. Show that $\mathbf{A}$ is PSD if and only if all of its eigenvalues are nonnegative.

Hint: Use the eigendecomposition of the matrix $\mathbf{A}$.

## 3 Probability Review

There are $n$ archers all shooting at the same target (bulls-eye) of radius 1. Let the score for a particular archer be defined to be the distance away from the center (the lower the score, the better, and 0 is the optimal score). Each archer's score is independent of the others, and is distributed uniformly between 0 and 1. What is the expected value of the worst (highest) score?

(a) Define a random variable $Z$ equal to the worst (highest) score, in terms of random variables that indicate each archer's score.

(b) Derive the Cumulative Distribution Function (CDF) of $Z$. *Hint: Recall the CDF of a random variable Z is given by* $F(z) = P(Z \leq z)$

(c) Let $X$ be a non-negative random variable. The Tail-Sum formula states that

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X \geq t)\, dt$$

Using both the Tail-Sum formula and the CDF of $Z$ you derived, calculate the expected value of $Z$. *Hint: Write* $\mathbb{P}(X \geq t)$ *in terms of the CDF of X.*

(d) Consider what happens to $\mathbb{E}[Z]$ as $n \to \infty$. Does this match your intuition?

 

# 4 Vector Calculus [1]

Below, $\mathbf{x} \in \mathbb{R}^d$ means that $\mathbf{x}$ is a $d \times 1$ column vector with real-valued entries. Likewise, $\mathbf{A} \in \mathbb{R}^{d\times d}$ means that $\mathbf{A}$ is a $d \times d$ matrix with real-valued entries. In this course, we will by convention consider vectors to be column vectors.

Consider $\mathbf{x}, \mathbf{w} \in \mathbb{R}^d$ and $\mathbf{A} \in \mathbb{R}^{d\times d}$. In the following questions, $\nabla_{\mathbf{x}}$ denotes the gradient with respect to $\mathbf{x}$, which, by convention, is a column vector.

Calculate the following derivatives.

(a) $\nabla_{\mathbf{x}}(\mathbf{w}^\top \mathbf{x})$

(b) $\nabla_{\mathbf{x}}(\mathbf{w}^\top \mathbf{A}\mathbf{x})$

(c) $\nabla_{\mathbf{A}}(\mathbf{w}^\top \mathbf{A}\mathbf{x})$

(d) $\nabla_{\mathbf{x}}(\mathbf{x}^\top \mathbf{A}\mathbf{x})$

(e) $\nabla^2_{\mathbf{x}}(\mathbf{x}^\top \mathbf{A}\mathbf{x})$

(f) Now let's apply our identities derived above to a practical problem. Given a design matrix $X \in \mathbb{R}^{nxd}$ and a label vector $Y \in \mathbb{R}^n$, the ordinary least squares regression problem is

$$w^* = \min_w \frac{1}{2}\|Xw - Y\|_2^2$$

Using parts (a)–(e), derive a necessary condition for $w^*$. *Note: We do not necessarily assume X is full rank! Hint: A necessary condition for a minimum point of a function is that it is a critical point, i.e. where the gradient is 0.*

[1] Good resources for matrix calculus are:

- The Matrix Cookbook: `https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf`
- Wikipedia: `https://en.wikipedia.org/wiki/Matrix_calculus`
- Khan Academy: `https://www.khanacademy.org/math/multivariable-calculus/multivariable-derivatives`
- YouTube: `https://www.youtube.com/playlist?list=PLSQl0a2vh4HC5feHa6Rc5c0wbRTx56nF7`.