1 Back to Basics: Linear Algebra

Let $X \in \mathbb{R}^{n\times m}$. We introduce some important terms and notation

The **Columnspace**, also called the range, or span, of X is $Range(X) := \{y \mid y = Xv\}$.
The **Rowspace** is $Row(X) := \{y \mid y = X^\top v\}$.
The **Nullspace**, or Kernel, of X is defined is $\mathcal{N}(x) := \{v \mid Xv = 0\}$.
The **Orthongal Complement** of a subspace, U, is a subspace, $U^\perp$ such that $u \in U, u' \in U^\perp \implies \langle u, u' \rangle = 0$

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

(a) Check the following facts:

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

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

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

(iv) $Row(X^\top X) = Range(X^\top X) = Row(X)$ *Hint: Use the relationship between nullspace and rowspace.*

(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 Rank nullity theorem says that for $X \in \mathbb{R}^{nxm}$ we have

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

Use the above results to prove this theorem. *Hint: The complementary subspace theorem says that for a vector space V and subspace U, we can always find a complementary subspace* $U^\perp$ *such that* $U + U^\perp = V$

2 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 that corresponds with the worst (highest) score.

(b) Derive the Cumulative Distribution Function (CDF) of 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 derived above, 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?

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

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

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

(f) Using parts (a) - (e), derive a necessary condition for w^*. *Note: We do not necessarily assume X is full rank!*

[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`.

