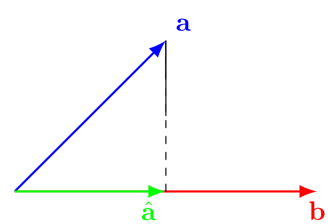
DSC 40A - Group Work Session 4
due Monday, Jan 29 at 11:59pm

Write your solutions to the following problems by either typing them up or handwriting them on another piece of paper. **One person** from each group should submit your solutions to Gradescope and **include all group members** so everyone gets credit.

This worksheet won't be graded on correctness, but rather on good-faith effort. Even if you don't solve any of the problems, you should include some explanation of what you thought about and discussed, so that you can get credit for spending time on the assignment.

In order to receive full credit, you must work in a group of two to four students for at least 50 minutes in your assigned discussion section. You can also self-organize a group and meet outside of discussion section for 80 percent credit. You may not do the groupwork alone.

1 Vector projection



The definition of the vector projection of a onto b when b is a unit vector:

$$\hat{a} = \frac{a^T b}{\|b\|} \frac{b}{\|b\|} = (a^T b) \cdot b$$

Show that $a - \hat{a}$ is orthogonal to b. This is why $\hat{a}$ is also known as the orthogonal projection.

Solution:

$$b^T(a - \hat{a}) = b^T a - b^T(a^T b)b = b^T a - (a^T b)b^T b = b^T a - a^T b = 0$$

Show that $\hat{a}$ is the closest vector in euclidean distance to a in the direction of b. That is, let $z = \lambda b, \lambda \in \mathbb{R}$, then the distance $\|a - z\|_2^2$ is minimized when $z = \hat{a}$. This follows the same rationale as homework 2 problem 4 in which you showed mean as the minimizer of mean squared loss.

Solution: For any vector z in the direction of b:

$$\|a - z\|_2^2 = \|a - \hat{a} + \hat{a} - z\|_2^2 = \|a - \hat{a}\|_2^2 + \|\hat{a} - z\|_2^2 + 2\underbrace{(a - \hat{a})^T(\hat{a} - z)}_{=0} \geq \|a - \hat{a}\|_2^2$$

2 Gradient with Respect to a Vector

The derivative of a scalar-valued function $f(\vec{w})$ with respect to a vector input $\vec{w} \in \mathbb{R}^n$ is called the gradient. The gradient of $f(\vec{w})$ with respect to $\vec{w}$, written $\nabla_{\vec{w}} f$ or $\frac{df}{d\vec{w}}$, is defined to be the vector of partial derivatives:

$$\frac{df}{d\vec{w}} = \begin{bmatrix} \frac{\partial f}{\partial w_1} \\ \frac{\partial f}{\partial w_2} \\ \vdots \\ \frac{\partial f}{\partial w_n}, \end{bmatrix}$$

where $w_1, \ldots, w_n$ are the components of the vector $\vec{w}$. In other words, the gradient of $f(\vec{w})$ with respect to $\vec{w}$ is the same as the gradient of $f(w_1, \ldots, w_n)$, a multivariable function of the components of $\vec{w}$.

Problem 1.

If $\vec{w} \in \mathbb{R}^n$, show that the gradient of $\vec{w}^T\vec{w}$ with respect to $\vec{w}$ is given by

$$\frac{d}{d\vec{w}}(\vec{w}^T\vec{w}) = 2\vec{w}$$

This should remind you of the familiar rule from single-variable calculus that says $\frac{d}{dx}(x^2) = 2x$.

Hints:

- First, start by writing $\vec{w}^T\vec{w}$ as a sum. What is the partial derivative of that sum with respect to w_1? With respect to w_2?
- If you're still confused, look at the example from Lecture 11 called "Example gradient calculation."

Solution: First, we rewrite $\vec{w}^T\vec{w}$ as a function of the components of $\vec{w}$, then we compute the partial derivatives to form the gradient vector. If

$$\vec{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix},$$

then

$$\vec{w}^T\vec{w} = w_1^2 + w_2^2 + \cdots + w_n^2 = \sum_{i=1}^{n} w_i^2.$$

Thinking of this as a function of multiple variables, the partial derivative with respect to w_i is just $2w_i$. That is,

$$\frac{\partial}{\partial w_i}(\vec{w}^T\vec{w}) = 2w_i.$$

Therefore, the gradient is

$$\frac{d}{d\vec{w}}(\vec{w}^T\vec{w}) = \begin{bmatrix} \frac{\partial}{\partial w_1}(\vec{w}^T\vec{w}) \\ \frac{\partial}{\partial w_2}(\vec{w}^T\vec{w}) \\ \vdots \\ \frac{\partial}{\partial w_n}(\vec{w}^T\vec{w}) \end{bmatrix} = \begin{bmatrix} 2w_1 \\ 2w_2 \\ \vdots \\ 2w_n \end{bmatrix} = 2\vec{w}.$$