CS 188
Summer 2023

# Discussion 6C

## 1 Vector Calculus

Let $\vec{x}, \vec{c} \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n \times n}$. For the following parts, before taking any derivatives, identify what the derivative looks like (is it a scalar, vector, or matrix?) and how we calculate each term in the derivative. Then carefully solve for an arbitrary entry of the derivative, then stack/arrange all of them to get the final result. Note that the convention we will use going forward is that vector derivatives of a scalar (with respect to a column vector) are expressed as a row vector, i.e. $\frac{\partial f}{\partial \vec{x}} = [\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, ..., \frac{\partial f}{\partial x_n}]$ since a row acting on a column gives a scalar. You may have seen alternative conventions before, but the important thing is that you need to understand the types of objects and how they map to the shapes of the multidimensional arrays we use to represent those types.

1. Show $\frac{\partial}{\partial \vec{x}}(\vec{x}^T\vec{c}) = \vec{c}^T$

2. Show $\frac{\partial}{\partial \vec{x}}||\vec{x}||_2^2 = 2\vec{x}^T$

3. Show $\frac{\partial}{\partial \vec{x}}(A\vec{x}) = A$

4. Show $\frac{\partial}{\partial \vec{x}}(\vec{x}^T A\vec{x}) = \vec{x}^T(A + A^T)$

5. Under what condition is the previous derivative equal to $2\vec{x}^T A$?

# 2 Solving Linear Regression with Vector Calculus

In this problem we will solve two variations of linear regression – ordinary least squares and ridge regression – using vector calculus.

1. *Ordinary Least Squares* Consider the equation $X\vec{w} = \vec{y}$, where $X \in \mathbb{R}^{n \times d}$ is a non-square data matrix, $w \in \mathbb{R}^d$ is a weight vector, and $y \in \mathbb{R}^n$ is vector of labels corresponding to the datapoints in each row of $X$.

   Consider the case where $n > d$, i.e. our data matrix $X$ has more rows than columns (tall matrix) and the system is overdetermined. **How do we find the weights $\vec{w}$ that minimizes the error between $X\vec{w}$ and $y$?** In other words, we want to solve $\min_{\vec{w}} \|X\vec{w} - \vec{y}\|^2$.

   **Use vector calculus to solve this optimization problem for $\vec{w}$.**

2. *Ridge Regression* Ridge regression can be understood as the unconstrained optimization problem

$$\arg\min_{\vec{w}} \|\vec{y} - X\vec{w}\|_2^2 + \lambda\|\vec{w}\|_2^2, \tag{1}$$

   where $X \in \mathbb{R}^{n \times d}$ is a data matrix, and $\vec{y} \in \mathbb{R}^n$ is the target vector of measurement values. What's new compared to the simple OLS problem is the addition of the $\lambda\|\vec{w}\|^2$ term, which can be interpreted as a "penalty" on the weights being too big.

   **Use vector calculus to expand the objective and solve this optimization problem for $\vec{w}$.**