# NYU Center for Data Science: DS-GA 1003

Machine Learning and Computational Statistics (Spring 2019)

Brett Bernstein

January 28, 2019

**Instructions**: Following most lab and lecture sections, we will be providing concept checks for review. Each concept check will:

- List the lab/lecture learning objectives. You will be responsible for mastering these objectives, and demonstrating mastery through homework assignments, exams (midterm and final), and on the final course project.
- Include concept check questions. These questions are intended to reinforce the lab/lectures, and help you master the learning objectives.

You are strongly encourage to complete all concept check questions, and to discuss these (and related) problems on Piazza and at office hours. However, problems marked with a ($\star$) are considered optional.

## Lab 1: Gradients and Directional Derivatives

### Multivariate Differentiation

#### Learning Objectives

1. Define the directional derivative, and use it to find a linear approximation to $f(\mathbf{x}+h\mathbf{u})$.
2. Define partial derivative and the gradient. Show how to compute an arbitrary directional derivative using the gradient.
3. For a differentiable function, give a linear approximation near a point $\mathbf{x}$ using the gradient.
4. Show that the gradient gives the direction of steepest ascent, and the negative gradient gives the direction of steepest descent.

**Concept Check Questions**

1. If $f'(x;u) < 0$ show that $f(x+hu) < f(x)$ for sufficiently small $h > 0$.

2. Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable, and assume that $\nabla f(x) \neq 0$. Prove

$$\operatorname*{arg\,max}_{\|u\|_2=1} f'(x;u) = \frac{\nabla f(x)}{\|\nabla f(x)\|_2} \quad \text{and} \quad \operatorname*{arg\,min}_{\|u\|_2=1} f'(x;u) = -\frac{\nabla f(x)}{\|\nabla f(x)\|_2}.$$

# Computing Gradients

**Learning Objectives**

1. Find the gradient of a function by computing each partial derivative separately.

2. Use the chain rule to perform gradient computations.

3. Compute the gradient of a differentiable function by determining the form of a general directional derivative.

**Concept Check Questions**

1. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by $f(x,y) = x^2 + 4xy + 3y^2$. Compute the gradient $\nabla f(x,y)$.

2. Compute the gradient of $f : \mathbb{R}^n \to \mathbb{R}$ where $f(x) = x^T A x$ and $A \in \mathbb{R}^{n\times n}$ is any matrix.

3. Compute the gradient of the quadratic function $f : \mathbb{R}^n \to \mathbb{R}$ given by

$$f(x) = b + c^T x + x^T A x,$$

   where $b \in \mathbb{R}$, $c \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n\times n}$.

4. Fix $s \in \mathbb{R}^n$ and consider $f(x) = (x-s)^T A (x-s)$ where $A \in \mathbb{R}^{n\times n}$. Compute the gradient of $f$.

5. Consider the ridge regression objective function

$$f(w) = \|Aw - y\|_2^2 + \lambda \|w\|_2^2,$$

   where $w \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$, $y \in \mathbb{R}^m$, and $\lambda \in \mathbb{R}_{\geq 0}$.

   (a) Compute the gradient of $f$.
   (b) Express $f$ in the form $f(w) = \|Bw - z\|_2^2$ for some choice of $B, z$. What do you notice about $B$?
   (c) Using either of the parts above, compute

$$\operatorname*{arg\,min}_{w\in\mathbb{R}^n} f(w).$$

6. Compute the gradient of

$$f(\theta) = \lambda\|\theta\|_2^2 + \sum_{i=1}^{n} \log(1 + \exp(-y_i\theta^T x_i)),$$

where $y_i \in \mathbb{R}$ and $\theta \in \mathbb{R}^m$ and $x_i \in \mathbb{R}^m$ for $i = 1, \ldots, n$.