Recitation 1

Gradients and Directional Derivatives

Brett Bernstein

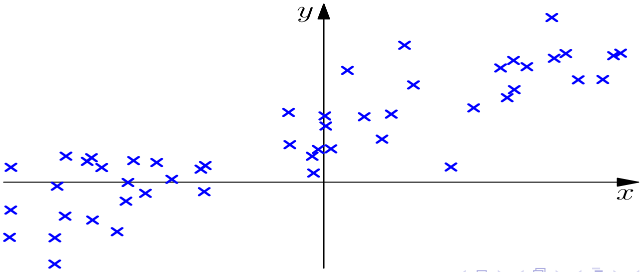
CDS at NYU

January 20, 2018

Intro Question

Question

We are given the data set $(x_1, y_1), \ldots, (x_n, y_n)$ where $x_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}$. We want to fit a linear function to this data by performing empirical risk minimization. More precisely, we are using the hypothesis space $\mathcal{F} = \{f(x) = w^T x \mid w \in \mathbb{R}^d\}$ and the loss function $\ell(a, y) = (a - y)^2$. Given an initial guess $\tilde{w}$ for the empirical risk minimizing parameter vector, how could we improve our guess?



Intro Solution

Solution

- The empirical risk is given by

$$\hat{R}_n(f) = \frac{1}{n}\sum_{i=1}^{n} \ell(f(x_i), y_i) = \frac{1}{n}\sum_{i=1}^{n}(w^T x_i - y_i)^2 = \frac{1}{n}\|Xw - y\|_2^2,$$

where $X \in \mathbb{R}^{n \times d}$ is the matrix whose ith row is given by x_i.

- Can improve a non-optimal guess $\tilde{w}$ by taking a small step in the direction of the negative gradient.

Single Variable Differentiation

- For $f : \mathbb{R} \to \mathbb{R}$ differentiable, the derivative is given by

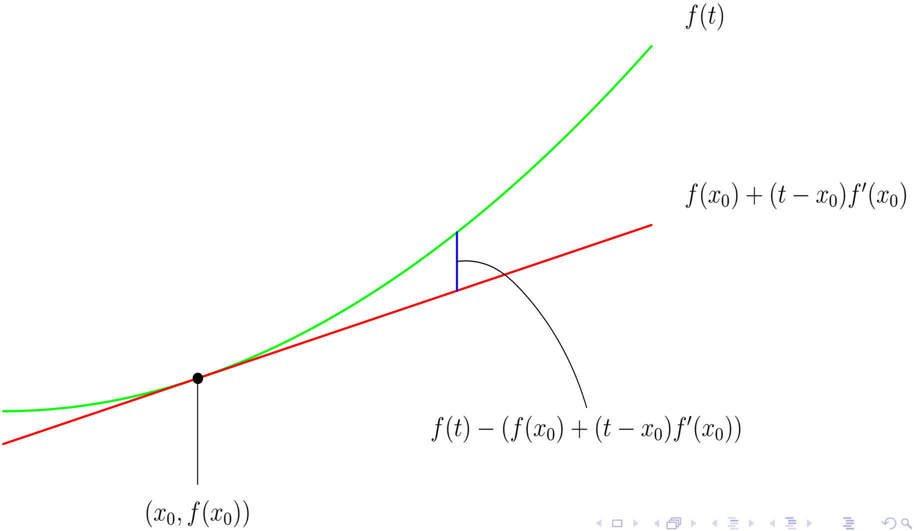
$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}.$$

- Can also be written as

$$f(x+h) = f(x) + hf'(x) + o(h) \quad \text{as } h \to 0,$$

  where $o(h)$ denotes a function $g(h)$ with $g(h)/h \to 0$ as $h \to 0$.
- Points with $f'(x) = 0$ are called *critical points*.

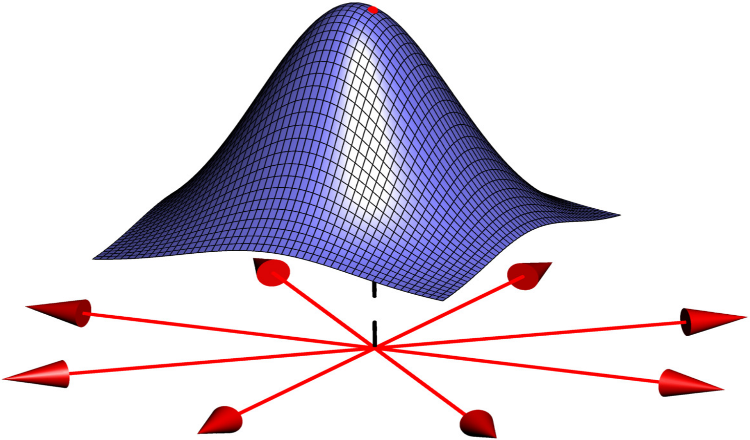
1D Linear Approximation By Derivative



Multivariable Differentiation

- Consider now a function $f : \mathbb{R}^n \to \mathbb{R}$ with inputs of the form $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$.
- Unlike the 1-dimensional case, we cannot assign a single number to the slope at a point since there are many directions we can move in.

Multiple Possible Directions for $f : \mathbb{R}^2 \to \mathbb{R}$



Directional Derivative

Definition

Let $f : \mathbb{R}^n \to \mathbb{R}$. The directional derivative $f'(x;u)$ of f at $x \in \mathbb{R}^n$ in the direction $u \in \mathbb{R}^n$ is given by

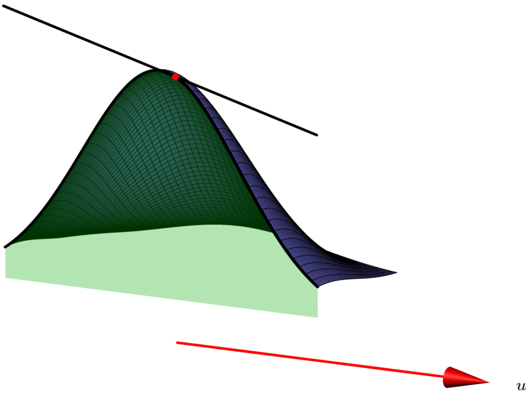
$$f'(x;u) = \lim_{h \to 0} \frac{f(x+hu) - f(x)}{h}.$$

- By fixing a direction u we turned our multidimensional problem into a 1-dimensional problem.
- Similar to 1-d we have

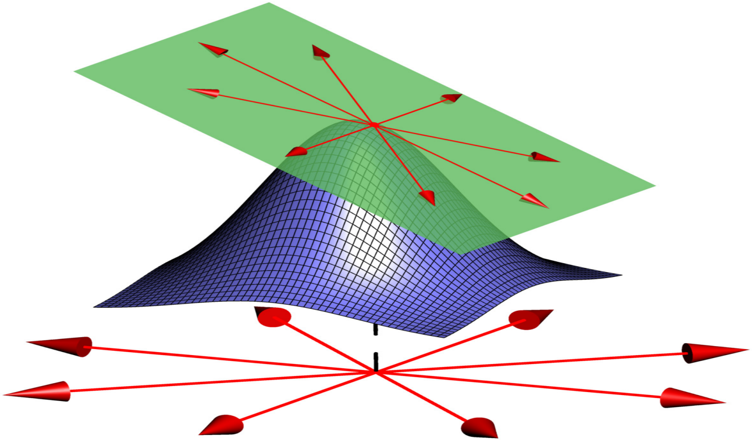
$$f(x+hu) = f(x) + hf'(x;u) + o(h).$$

- We say that u is a *descent direction* of f at x if $f'(x;u) < 0$.

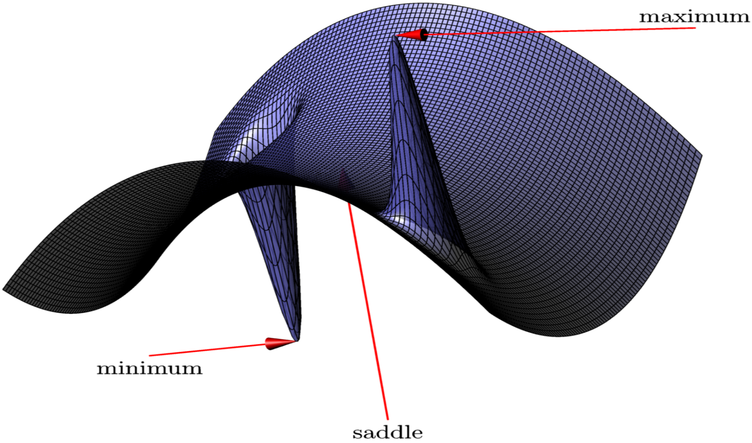
Directional Derivative as a Slope of a Slice



Tangent Plane for $f : \mathbb{R}^2 \to \mathbb{R}$



Critical Points of $f : \mathbb{R}^2 \to \mathbb{R}$



Computing Gradients

Question

For each of the following functions, compute the gradient.

1. $f : \mathbb{R}^3 \to \mathbb{R}$ is given by

$$f(x_1, x_2, x_3) = \log(1 + e^{x_1+2x_2+3x_3}).$$

2. $f : \mathbb{R}^n \to \mathbb{R}$ is given by

$$f(x) = \|Ax - y\|_2^2 = (Ax - y)^T(Ax - y) = x^T A^T Ax - 2y^T Ax + y^T y,$$

for some $A \in \mathbb{R}^{m \times n}$ and $y \in \mathbb{R}^m$.