Recitation 1

Gradients and Directional Derivatives

Brett Bernstein

CDS at NYU

January 25, 2017
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?
1D Linear Approximation By Derivative

\[ f(t) - (f(x_0) + (t - x_0)f'(x_0)) \]

\[ (x_0, f(x_0)) \]

\[ f(x_0) + (t - x_0)f'(x_0) \]
Multiple Possible Directions for $f : \mathbb{R}^2 \to \mathbb{R}$
Directional Derivative as a Slope of a Slice

\[ u \]
Tangent Plane for $f : \mathbb{R}^2 \rightarrow \mathbb{R}$
Critical Points of $f : \mathbb{R}^2 \rightarrow \mathbb{R}$
Question

For each of the following functions, compute the gradient.

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

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

2. $f : \mathbb{R}^n \rightarrow \mathbb{R}$ 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 $y \in \mathbb{R}^m.$