50.021 – AI

Alex

Week 02: Gradients, linear hyperplanes

[The following notes are compiled from various sources such as textbooks, lecture materials, Web resources and are shared for academic purposes only, intended for use by students registered for a specific course. In the interest of brevity, every source is not cited. The compiler of these notes gratefully acknowledges all such sources.]

Due: week3 Thursday, 6th of June, 6pm

1 Disastrous Derivatives

Compute the directional derivative $Df(X)[H]$ in direction H for:

$$\begin{aligned} f(X) &= aX, && a \in \mathbb{R}^{1\times d}, X \in \mathbb{R}^{d\times k} \\ f(X) &= XX^\top, && X \in \mathbb{R}^{d\times n} \end{aligned}$$

What will be the shape of the direction H in $Df(X)[H]$ for these two? Is it a real number, a vector or a matrix? Express it as $\mathbb{R}^{1\times 1}$ if you think it will be a scalar, as $\mathbb{R}^{d\times 1}$ if you think it is a vector, or as $\mathbb{R}^{d\times e}$ if you think it is a matrix. What will be $Df(X)[H]$? Hint: you can write it as product of matrices if you like it (instead of summing in the flavor of $\sum_{ijk} c_{ijk}$).
Compute the directional derivative $Df(X)[H]$ in direction H for:

$$\begin{aligned} f(X) &= XBX, && X \in \mathbb{R}^{d\times d} \\ f(X) &= AXBX^\top CX, && X \in \mathbb{R}^{d\times d}, \{A, B, C\} \in \mathbb{R}^{d\times d} \end{aligned}$$

Hint: remember in class $f(x) = A(x)C(x)$? There we noted that we have the structure $f(X) = B(A(x), C(x))$ where B is a bilinear mapping. For a bilinear mapping the derivative was a sum of two terms, for a trilinear mapping the sum was of three terms.

Can something similar be done with a linear, a bilinear or a trilinear function here?

Compute the directional derivative $Df(X)[H]$ in direction H for:

$$f(X) = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 1 & x_2^2 \\ \ln x_1 & 3 \end{pmatrix}$$

2 Sizzling SVMs

Some people try to solve SVM-like criteria with SGD solvers. That makes sense when the number of samples is very high and you need the speed (check vowpal wabbit for example)

$$L = 0.5\|w\|^2 + C\sum_i \max(0, 1 - y_i(w \cdot x_i + b))$$

Compute the gradient of L with respect to w. Hint: write the max as something depending on indicator functions

$$\begin{aligned} \max(0, x) &= \begin{cases} x & x > 0 \\ 0 & x \leq 0 \end{cases} \\ &= 1[x > 0]x \; (+1[x < 0]0) \end{aligned}$$

3 Haunting Hyperplanes

Suppose you have in 2 dimensional space the following dataset: $D_n = \{(x_1 = (1,0), y_1 = +1), (x_2 = (0,1), y_2 = +1), (x_3 = (1,1), y_3 = -1)\}$. Find a linear classifier $f(x) = wx + b$ (means its parameters) which predicts all points correctly.

Suppose you have in 3 dimensional space the following dataset: $D_n = \{(x_1 = (1,0,1), y_1 = +1), (x_2 = (-1,-3,2), y_2 = -1)\}$. Find a linear classifier $f(x) = wx + b$ (means its parameters) which predicts all points correctly.

Hint: draw it. matplotlib if you never felt too cozy with studying arts.