50.039 – Theory and Practice of Deep learning

Alex

Week 02

Due: week3 wednesday, 6pm

1 Disastrous Derivatives

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

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

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:

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

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:

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

2 n-dim Hyperplanes are a piece of bunny!

Here I want you to think a bit geometrically about these hyperplanes without the need to draw them.

2.1 3 dims

Suppose you have in 3 dimensional space the following dataset: $D_n = \{(x_1 = (-5, 1, 2), y_1 = +1), (x_2 = (1, 2, -3), y_2 = -1)\}$.
Find a linear classifier $f(x) = wx + b$ (means its parameters) which predicts all points correctly.

Hint:
It can help to think of a hyperplane defined in terms of an orthogonal vector and a bias term. Get the orthogonal vector to the hyperplane right first, bias you can determine then afterwards by plugging in points, and solving for the bias.
For two points, what would be a good orthogonal vector for a hyperplane if you have two points which need to be separated??

2.2 5 dims I

Suppose you have in 5 dimensional space the following dataset: $D_n = \{(x_1 = (1, 0, 1, -2, 6), y_1 = +1), (x_2 = (3, 1, -3, 5, 1), y_2 = -1)\}$.
Its the same thing as above, now in 5 dims. Find a linear classifier $f(x) = wx+b$ (means its parameters) which predicts all points correctly.

2.3 5 dims II

Suppose you have in 5 dimensional space the following dataset: $D_n = \{(x_1 = (1, 0, 1, 6, 3), y_1 = +1), (x_2 = (3, 1, -3, 5, 1), y_2 = -1), (x_3 = (2, -1, 0, 4, 2), y_2 = -1)\}$.
Good news: you cant draw it but still solve it. An alternative way is to project these three points into a 2-dim subspace, by the way, and solve it in the subspace.

How to solve it:

You can draw a line between the two points x_2, x_3 of the same label. This line can be part of many hyperplanes (its a 4d space of all possible hyperplanes). The vector $x_2 - x_3$ is parallel to this line .
These are the steps:

I **Now choose a vector w:**

- which is **not** parallel to $x_2 - x_3$,
- which is **not** orthogonal to $x_1 - x_3$,

and which will serve as candidate for the orthogonal of the hyperplane to be found.

II **Answer:**

- what is a possible mathematical criterion to test that w is not parallel to the line $x_2 - x_3$?
- what is a possible mathematical criterion to test that w is not orthogonal to the line $x_1 - x_3$?

III Next, **make this vector w into a vector w_2 which is orthogonal** to the line $z = x_2 - x_3$ by using the following:

$$w_2 = w - (w \cdot z)\frac{z}{\|z\|^2}$$

Why do I want you to do it?

$$w_2 \cdot (x_2 - x_3) = 0$$

implies that:

$$w_2 \cdot x_2 + b = w_2 \cdot x_3 + b$$

means $f(x_2) = f(x_3)$ no matter what choice of bias. Once you choose a bias, both points will lie on the same side of the hyperplane! Also note that the projection formula has a more symmetrical writing form which makes clearer what it does:

$$w_2 = w - (w \cdot \frac{z}{\|z\|})\frac{z}{\|z\|}$$

We remove from w the component in the direction of $\frac{z}{\|z\|}$. The amount of the component is $w \cdot \frac{z}{\|z\|}$

IV **Show by calculation** that w_2 is orthogonal to $z = x_2 - x_3$. You can do this with symbols, without plugging in numbers.

V **Use w_2 to for your classifier, and determine a bias b which separates points.**

VI Answer: If w would be parallel to $x_2 - x_3$, why in some case it could become unsuitable as a separating hyperplane ? Think of it in 2 or 3 dimensions to produce an answer.

Off the tasks: This gives you a hint, why I asked you in III. to remove the component in w parallel to $x_2 - x_3$ before solving for a bias.

Offtopic: this works for any K points z_k in D dims $(K <= D)$ which are linearly separable from one bad point x.

- find a hyperplane spanning these K points z_k – for this you can consider the span of $z_2 - z_1, z_3 - z_1, \ldots, z_k - z_1$.
- find a vector w_2 which is orthogonal to the hyperplane ($D - K + 1$ dim space, it must be orthogonal to all the $z_i - z_1$), and which is not orthogonal to at least one of the $z_k - x$.
- use that w_2 to solve for the bias.