w01-functions-1

Magnus Chr. Hvidtfeldt

Technical University of Denmark, Lyngby, DK,

s255792@dtu.dk

Category: Lecture notes for 01002-mat1b week 1, functions with several variables, norm, inner product spaces, continuity.

1 Vector function with several variables

The core idea of this week is a vector function with several variables. We have

$$f : \mathrm{dom}(f) \to \mathbb{R}^k \tag{1.1}$$

Where $\mathrm{dom}(f) \subset \mathbb{R}^n$.

It sort of acts like a parametric equation, where we have

$$f(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \cdots \\ f_k(\mathbf{x}) \end{bmatrix} \tag{1.2}$$

For $\mathbf{x} = [x_1, x_2, \ldots, x_k]^T \in \mathbb{R}^k$. Thus f is evaluated at each point in the vector x. This means that a single evaluation $f_i : \mathrm{dom}(f) \to \mathbb{R}$.

1.1 Surjective and injective

We can evaluate surjective to see if $\mathrm{codom}(f) = \mathrm{Im}(f)$. Likewise, we can evaluate injective, check if $\dim \ker(A) = \{0\}$.

2 Norm and inner product spaces

If we want to find the magnitude of a vector, we can take the regular norm as such

$$\|\mathbf{x}\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} \tag{2.1}$$

This is for a vector $\mathbf{x} \in \mathbb{R}^n$.

2.1 p-norm

If we want to extend the norm to complex numbers, given a vector $\mathbf{x} \in \mathbb{C}^n$, and for generally other p's, we have

$$\|\mathbf{x}\|_p = \left(\sum_{k=1}^{n} |x_k|^p \right)^{1/p} \tag{2.2}$$

Which means we have for important norms, what we consider the regular norm unless otherwise stated is $\|\mathbf{x}\|_2$. We also have the ∞-norm defined as

$$\|\mathbf{x}\|_\infty = \max_{k=1,..,n} |x_k| \tag{2.3}$$

For a vector $\mathbf{x}$.

2.2 Inner product spaces

We define the inner product space, given two vectors $\mathbf{x}$ and $\mathbf{y}$, as the dot product, extending to $\mathbb{C}^n$, we have

$$\langle \mathbf{x}, \mathbf{y} \rangle = x_1\bar{y_1} + x_2\bar{y_2} + \cdots + x_n\bar{y_n} \tag{2.4}$$

Where if either $\mathbf{x}$ or $\mathbf{y}$ is $\mathbb{R}$, it will simply be the dot product. But if it's complex, it will be the complex conjugate of $\mathbf{y}$.

3 ReLU and neural network

3.1 ReLU

ReLU is a non-linear activation function, that makes a neural network be non-linear, allowing the network to have millions of layers and activate the ideas we need. It is quite simple to differentiate,

$$\mathrm{Re}LU(x) = \begin{cases} x, x \geq 0 \\ 0, x < 0 \end{cases} \tag{3.1}$$

And thus for a network, vector, of millions of entries, we have

$$\mathbf{ReLU}(\mathbf{x}) = \begin{bmatrix} ReLU(x_1) \\ ReLU(x_2) \\ \cdots \\ ReLU(x_k) \end{bmatrix} \tag{3.2}$$

3.2 Neural network

To design a neural network we have

$$\Phi(x) = A_1x + b_1 \tag{3.3}$$

Called an affine transformation, if we have several layers we have

$$\Phi(x) = A_1\sigma_1(A_2x + b_2) + b_1 \tag{3.4}$$

Where σ_1 is our activation function (like ReLU or softmax), A is our matrix of weights (like in MNIST, the weights correspond to the hues 0-255 of the number, which is then flattened from $\mathbb{R}^{28\times 28} \to \mathbb{R}^{784}$) and b is our bias vector.

The output $\Phi(x)$ is, in the case of MNIST, a probability vector in $\mathbb{R}^{10}$ of how likely it is that it is the number x_i.

4 Level sets (niveaukurver)

Take a function of several variables, if we take a cut through the horizontal line, we get a 2D curve, which is the level set defined as follows

$$\{x \in \mathrm{dom}(f) \mid f(x) = c\} \tag{4.1}$$

For some c.

5 Continuity definition

If something is continuous, we check that $x \to x_0 \implies f(x) \to f(x_0)$. More simply, we have to have $\lim_{x \to x_0} f(x) = f(x_0)$ for some arbitrary x_0, then it will be true for any a.

Let $\varepsilon > 0$ be given. Then there exists a $\delta > 0$ such that, if $|x - x_0| < \delta$, then we have that $|f(x) - f(x_0) < \varepsilon$.

Written formally,

$$\forall \varepsilon > 0, \exists \delta > 0 : 0 < |x - x_0| < \delta \implies |f(x) - f(x_0)| < \varepsilon \tag{5.1}$$

Where x_0 is some arbitrary point.

Thus, to prove something is continuous, we simply have to prove that the limit $\lim_{x \to x_0} f(x) = f(x_0)$ exists for some arbitrary x_0.

5.1 Example

Prove that $f(x) = ax + b$ is continuous.

Let $\varepsilon > 0$ be given. Then we have $|x - x_0| < \delta$, furthermore

$$\begin{aligned} |ax + b - (ax_0 + b)| = |ax - ax_0| &< \varepsilon \qquad (5.2) \\ \to |a||x - x_0| &< \varepsilon \qquad (5.3) \\ \to |x - x_0| &< \frac{\varepsilon}{|a|} \qquad (5.4) \end{aligned}$$

Choose $\delta = \left\lceil \frac{\varepsilon}{|a|} \right\rceil$.

Then, for all $n > \delta$, it follows that

$$|f(x) - f(x_0)| = |a||x - x_0| < |a|\frac{\varepsilon}{|a|} = \varepsilon \tag{5.5}$$

Which proves that the limit holds for x_0. Since x_0 was arbitrary, we have shown that f is continuous. □

6 Partial derivatives and gradient

We have $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ for finding derivates. Where, with respect to x, we set y to be a constant and differentiate like normal. Similarly, with respect to y, we set x to be a constant and differentiate like normal.

6.1 Gradient

We have

$$\nabla(x, y) = \begin{bmatrix} \frac{\partial f}{\partial x} f(x, y) \\ \frac{\partial f}{\partial y} f(x, y) \end{bmatrix} \tag{6.1}$$

Visually, the gradient represents, at a given point, where the steepest descend is.?