Introduction to Neural Networks

Johns Hopkins University
Engineering for Professionals Program

605-447.71/625-438.71

Dr. Mark Fleischer

Module 2.4: Mathematical Review-Metric Spaces

This Sub-Module Covers …

- Some mathematical review of Metric Spaces and will set the stage for using:
  - directional derivatives in conjunction with calculus based optimization to define the
  - Method of Steepest Descent (MOSD)---used to train Perceptrons.
- Also provides additional insights into dynamical systems.

What is ‘Distance’ and ‘Magnitude’?

- Need capability to rank and/compare objects using a simple criterion.
- Want a flexible yet abstract notion of distance or length or magnitude.
- The following principles provide a very abstract, general way of defining the essential properties of a ‘length’.

Length

Desirable properties of a "length"(magnitude or norm):

1. Positivity: $||\mathbf{y}|| \geq 0$ for all y and $||\mathbf{y}|| = 0$ if and only if $\mathbf{y} = \mathbf{0}$.
2. Homogeneity: $||c\mathbf{y}|| = |c|\ ||\mathbf{y}||$ for all scalars c and vectors $\mathbf{y}$.
3. The Triangle Inequality: $||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$ for all vectors $\mathbf{x}$ and $\mathbf{y}$.

Examples of a Norm

$$|\vec{a}| \equiv \|\vec{a}\| = \left[\sum_{i=1}^{n} a_i^2\right]^{1/2}$$

This is called the Euclidean Norm.

$$|\vec{a}|_p \equiv \|\vec{a}\|_p = \left[\sum_{i=1}^{n} a_i^p\right]^{1/p}$$

This is the *p*-norm.

If $\mathbf{y} = \sum_{1=1}^{n} \alpha_i \mathbf{v}_i$ for some set of *n* linearly independent vectors $\mathbf{v}_i$. Then

$$\|\mathbf{y}\|_{\mathbf{V}} \equiv \sum_{i=1}^{n} |\alpha_i|$$

is a norm with respect to the matrix $(\mathbf{v}_1 | \mathbf{v}_2 | \cdots | \mathrm{v}_n)$.

Metric Spaces

We've already alluded to some issues for measuring mathematical things.
We need tools that allow us to *rank* mathematical objects and *rank relationships* between and among mathematical objects.

1. $\rho(x, y) = \rho(y, x) \geq 0.$ Positivity
2. $\rho(x, x) = 0; \quad \rho(x, y) = 0 \Leftrightarrow x = y.$ Homogeneity
3. $\rho(x, z) + \rho(z, y) \geq \rho(x, y).$ Triangle Inequality

where ρ is a function defined on $M \times M$, *hence* (M, ρ) is called a Metric Space

The metric function ρ can take on many forms!

We define ρ based on what is necessary and convenient!

Various Forms of Metrics

$$\rho_\infty(x,y) = \operatorname*{Max}_i \left|x_i - y_i\right|$$

$$\left|x_i - y_i\right| \le \left|x_i - z_i\right| + \left|z_i - y_i\right|$$

$$\rho\big(f(x), g(x)\big) = \operatorname*{Max}_{x\in[a,b]} \left|f(x) - g(x)\right|$$

$$\rho(\vec{a},\vec{b}) = \left\|\vec{a} - \vec{b}\right\| = \left[\sum_{i=1}^{n} (a_i - b_i)^2\right]^{1/2} = \left(\vec{a} - \vec{b}, \vec{a} - \vec{b}\right)^{1/2}$$

Orthogonality and Angles

We can posit that for any two mathematical objects, there corresponds a number between -1 and 1 that conveys how these objects are related.

$$\cos\theta = \frac{\langle \mathbf{a}, \mathbf{b} \rangle}{\|\mathbf{a}\| \|\mathbf{b}\|}$$

We can say that this corresponds to the angle between two vectors.

Convergent Sequences

Limit Points: A sequence $\{x_n\}$ is convergent and has a limit point

$$x^* \in M, \text{ if } \rho(x_n, x^*) \rightarrow 0 \text{ as } n \rightarrow \infty,$$

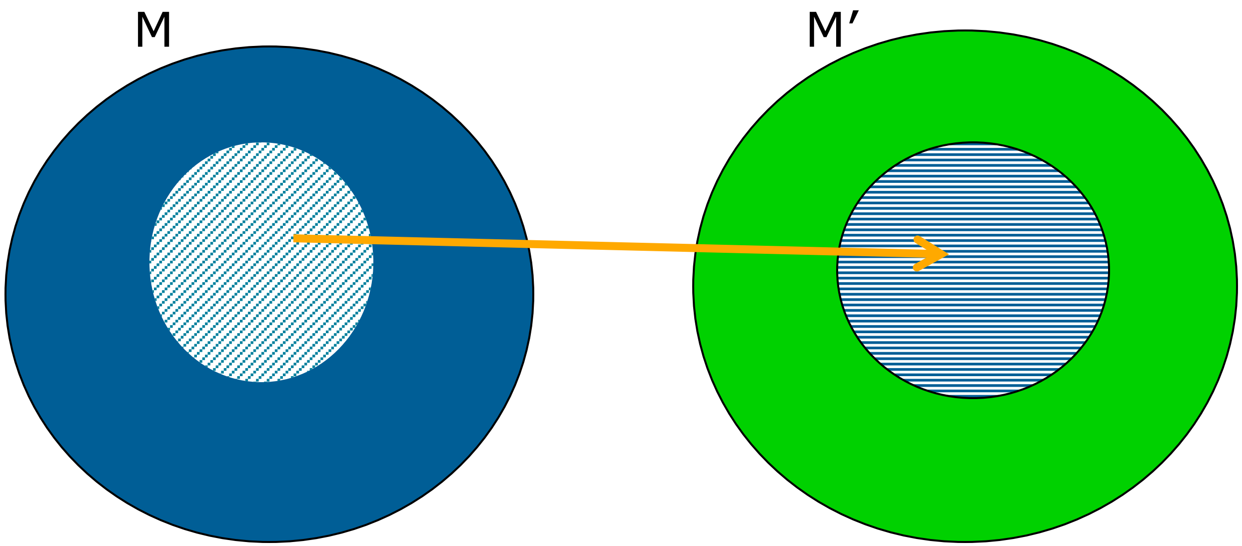
$$x_n \rightarrow x^*$$

Mappings in Metric Spaces

- Define two sets and two Metric Spaces:

$$D(f) = \{x : x \in M \wedge f(x) \text{ is defined}\}$$

$$R(f) = \{x' : x' \in M' \wedge \exists\, x \in D(f) \ni x' = f(x)\}$$



Boundedness

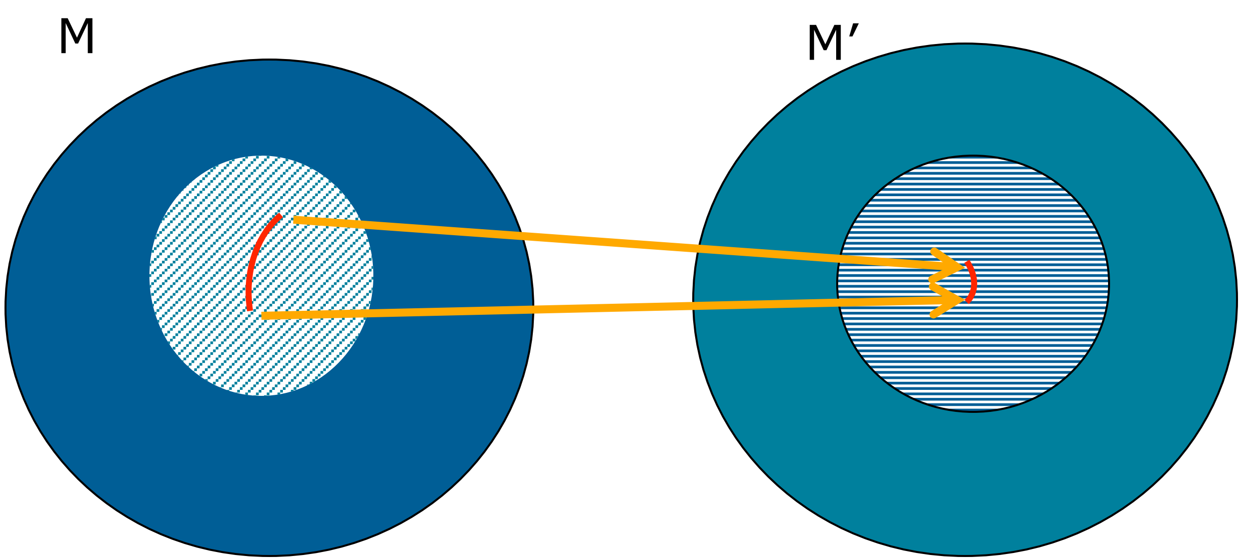
$$\exists k \geq 0, \ni \forall x, y \in D(f)$$

$$\underbrace{\rho'\big(f(x), f(y)\big)}_{\text{Metric value associated with the range.}} \leq k\underbrace{\rho(x, y)}_{\text{Metric value associated with the domain.}}$$

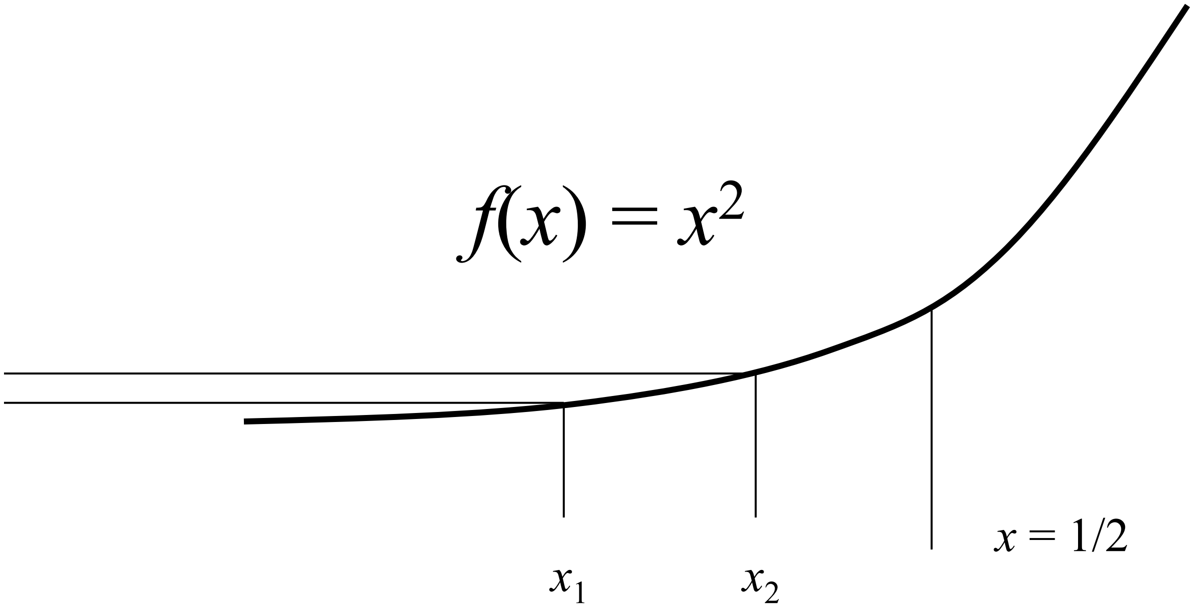
Contraction Mapping

A Mapping is a contraction if it is bounded and

$$\exists k \geq 0 \ \wedge \ 0 \leq k < 1, \ni \forall x, y \in D(f)$$
$$\rho'\big(f(x), f(y)\big) \leq k\rho(x, y)$$



An Example



Define $\rho(x_1, x_2) = |x_1 - x_2|$

A Useful Fact

$\Delta y = f'(\xi)\Delta x$ where ξ is some "intermediate" value.

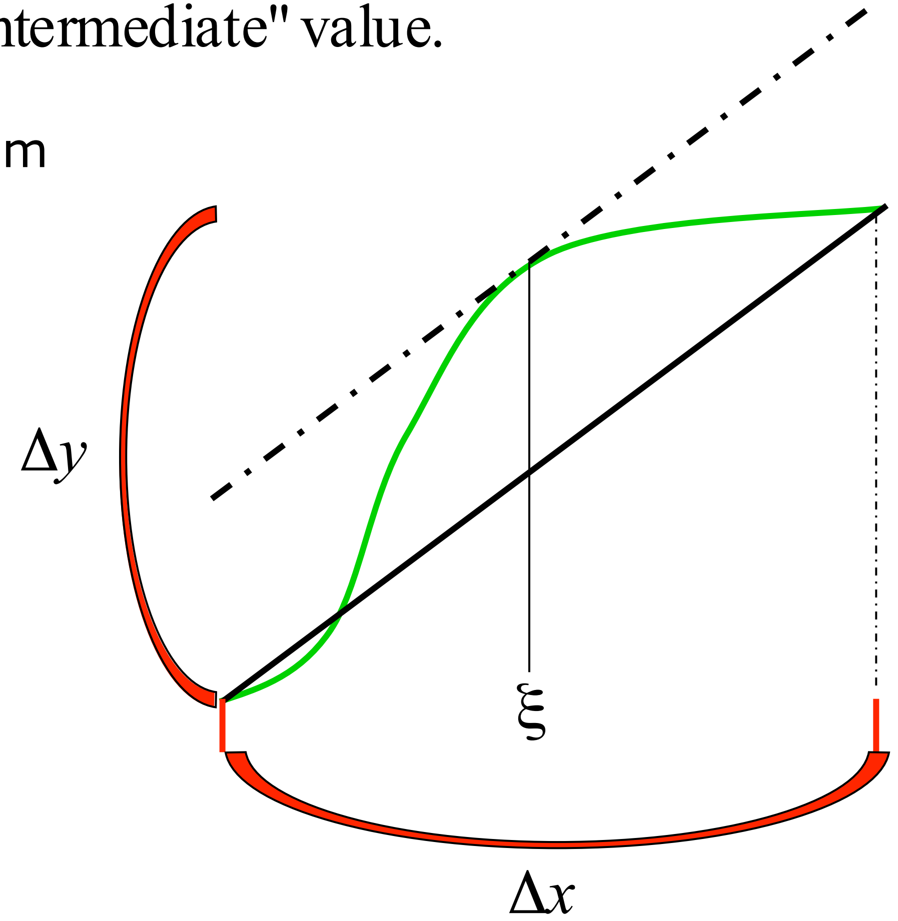
The Mean Value Theorem

$$|\Delta y| = |f(x_1) - f(x_2)|$$

$$= |f'(\xi)\Delta x|$$

$$\leq |f'(\xi)||\Delta x|$$

$$= f'(\xi)|x_1 - x_2|$$



Putting It All Together

$$|\Delta y| = |f(x_1) - f(x_2)| = \rho(f(x_1), f(x_2))$$

Therefore,

$$\rho(f(x_1), f(x_2)) \leq f'(\xi)\rho(x_1, x_2)$$

and since $0 \leq x_1, x_2 < \frac{1}{2}$, and $f'(x) = 2x$

then it follows that $f'(\xi) < 1$

$$\forall x, y \in D(f) \text{ and } 0 \leq k < 1$$

$$\rho'(f(x), f(y)) \leq k\rho(x, y)$$

is a Contraction Mapping

A Numerical Example

$$\rho\left(\frac{1}{4},\frac{1}{2}\right)=\frac{1}{4}$$

$$\begin{aligned}\rho\left(f\left(\tfrac{1}{4}\right),f\left(\tfrac{1}{2}\right)\right)&=\rho\left(\left(\tfrac{1}{4}\right)^2,\left(\tfrac{1}{2}\right)^2\right)\\&=\rho\left(\tfrac{1}{16},\tfrac{1}{4}\right)\\&=\tfrac{3}{16}\end{aligned}$$

Note that $\frac{3}{16}<\frac{1}{4}$

How Would You Use This in The Context of Fixed Points?

$$\forall x_1, x_2 \in D(f) \text{ and } 0 \le k < 1$$
$$\rho'\big(f(x_1), f(x_2)\big) \le k\rho(x_1, x_2)$$

$$\forall x_1, x_2 \in D(f) \text{ and } 0 \le k < 1$$
$$\rho'\big(x^*, f(x_2)\big) \le k\rho(x^*, x_2)$$

because by definition, a fixed point x^* is such that

$$f\left(x^*\right) = x^*$$

Dynamical Systems

- Neural Networks can be fashioned into dynamical systems.
- Feeding outputs back as inputs and cycling them ala fixed point exercise.
- How does such a system behave?
- Stay tuned.