Statistical Machine Learning
Reading Assignment 1 Report

Alireza Sadeghi - Mohsen Shojaee

March 1, 2016

1 A Taste of Real Analysis

Real Analysis is the field of mathematics on which probability theory is founded, it is therefore convinient to first introduce some basic concepts from real analysis. We begin the this chapter with the definiation of metric space and use the notion of metric to define two core concepts: convergence and continuity. In 1.1.2 we revie the natural topoly defined on a metric space based on the metric function and then in 1.2 we provide a more general view of topological spaces.

1.1 Metric Space

A metric space is a set M together with a metric $d : M \times M \to \mathcal{R}$ satisfying following four properties

- d is non-negative
- $d(x, y) = 0$ iff $x = y$
- Symmetry: $d(x, y) = d(y, x)$
- Triangle Inequality: $d(x, z) \leq d(x, y) + d(y, z)$

Strictly speaking, the pair (M, d) is the metric space as different metric functions can be defined on same M, consider for example $\mathcal{R}^n$; a well-known class of metrics defined on $\mathcal{R}^n$ is **Minkowski Norm**:

$$d_p(x, y) = \big(\sum_{i=1}^{n}(x_i - y_i)^p\big)^{1/p}$$

which for all values of $p \geq 1$ is a valid metric function.

1.1.1 Convergence & Continuity

There are different ways for defining convergence, we follow [?] and use sequence/subsequence approach. A sequence (p_n) is a list of points $p_1, p_2, \ldots$ in M. Formaly, a sequence is a function $f : \mathbb{N} \to M$ in which $f(n) = p_n$. The sequence (p_n) **converges to the limit** p in M (and denote this by $(p_n) \to p$ if:

$$\begin{aligned} &\forall \epsilon > 0 \quad \exists N \in \mathbb{N} \quad \text{such that} \\ &n \geq N \Rightarrow d(p_n, p) < \epsilon \end{aligned}$$

Having defined convergence, contiuity can be described: For a function $f : M \to N$ between two metric spaces (M, d_M) and (N, d_N), we say that function is continous if it preserves sequential convergence, that is if $(p_n) \to p$ then $(f(p_n)) \to f(p)$.

The sequence definitation of continiutity stated above is equivalent with the more familiar definiation using (ϵ, δ) condition:

Theorem 1 *$f : M \to N$ is continouts if and only if for each $\epsilon > 0$ and $p \in M$ there exisits $\delta > 0$ such $\forall x \in M :\quad d_M(x, p) < \delta \Rightarrow d_N(f(x), f(p)) < \epsilon$*

1.1.2 Topology of Metric Space

Althogh topology can be defined on non-metric spaces (as we will do so in the next section) there is a *natural* topology induced on metric spaces induced by the distance function. To this end we need to define notion of **openness** and **closeness** based metric and convergences. We say that point $p \in M$ is a limit of $S \subset M$ if there exists a sequence in S like (p_n) that $(p_n) \to p$

Closeness: S is a closed set if it contains all it limits. **Openness:** S is an open set if for each $p \in S$, $r > 0$ exists such that

$$d(p, q) < r \Rightarrow q \in S.$$

that is for each point in S an small ball around it is also in S.

One can simply prove that complement of an open set is closed and vice versa. However (like doors) sets can be neighter open nor closed and unlike doors they can be both at the same time.

Theorem 2 *Now the collection $\mathcal{T}$ of all open sets of M is the topolgy of M, i.e. is satisfies the following three properties:*

- $M, \Phi \in \mathcal{T}$
- *The intersection of finitely many open sets is an open set*
- *The union of arbitrarily many open sets is an open set.*

1.2 Topology: a More General Perspective

The three properties stated in 2 are the definiation of topology, one can *handcraft* a collection $\mathcal{T}$ that satisfies these properties and call it the collection of open sets of M, even if they does not satisfy the definiation of openness based on metric or even M is not a metric space at all.

2 Theory of Probability

2.1 Introduction

In many cases where statistics and statistical inference is an essential component of situation analysis, one encounters many discrete and continuous random variables and vectors and matrices. These are all special cases of a more general type of random quantity. The generalization of these notions to random quantities is through a notion of *measure*.

2.2 Measure Theory

Measure, to be defined shortly is a way of assigning numerical values to the sizes of sets. Since it's used to give sizes to sets, it's domain is a collection of sets.

In order to define this "collection of sets" notion more thoroughly, we call a collection of sets that is closed under taking complements and finite unions, a ***field***.

A field, that is closed under taking countable unions is called a ***σ-field***.

A σ-field that is generated by the collection C of open subsets of a topological space is called a ***Borel σ-field***

2.3 Measurable Functions

Suppose S is a set with a σ-field A of subsets, and let T be another set with a σ-field C of subsets. Now consider a function $f : S \to T$. We say f is *measurable* if for every B $\in$ C, $f^{-1} \in A$.

If f is measurable, one-to-one and onto, and $f^{-}1$ is also measurable, we say that f is *bimeasurable*, and also in another case, if the two sets *S, T* are topological spaces with Borel σ-fields, a measurable function is *Borel measurable.*

2.4 Mathematical Probability

If we'd want to define probability with a measure theoretic approach, we'd say that a measure space *(S, A, μ)* is a probability space and μ is a probability if $\mu(S) = 1$. We call each element of A an *event*, and a measurable function X from S to some other space *X, B* is called a random quantity.

When X is $\Re$ with the *Borel σ-field*, this random quantity is called a *Random Variable.* The probability measure μ_x induced on *(X, B* by X from is called the *distribution* of X. The expected value or mean

3 Stochastic Processes

3.1 Definition

A stochastic process X$_t$ t $\in T$ is a collection of random variables X$_t, takingvaluesinacommonmeasurespace(S, X$

This definition means that for each $t \in T$, $X_t(\omega)$ is an S $\rightarrow X{-}measurable function from \Omega$ to S, which yields a probability on S.

Some examples of stochastic processes are as follows:

- Every random variable is a trivial stochastic process.
- Let $T = 1, 2, \cdots, k$ and $S = \Re$. Then X_tt $\in T$ is a random vector in $\Re^k$.
- One-sided random sequence, in which we define T to be 1, 2, $\cdots . Two-sided random sequences, in which we define T to be \mathbb{Z}.$
- Spatially-discrete random fields, in which we let T to be $\mathcal{Z}$.
- and some other types.

3.2 Random Functions

$X(t, \omega)$ has two arguments, t and ω. If we fix the value of t, $X_t(\omega$ is a random variable. For each fixed value of ω though, X(t) is a function from T to S, we call it a random function.