BRAJ KISHOR PRASAD, brajki@rediffmail.com, Department of MCA, 4th Semester
MCA DSE4T2: *Introduction to Machine Learning*

Space, Instance Space, Concept Space and Hypothesis Space

Space:

In mathematics, a **space** is a set (sometimes called a universe) with some added structure.

While modern mathematics uses many types of spaces, such as Euclidean spaces, linear spaces, topological spaces, Hilbert spaces, or probability spaces, it does not define the notion of "space" itself

Instance Space

This is associated within a field of math called computational learning theory (CLT). CLT is inherently abstract since it's trying to derive general observation about unknown target functions. That's why we get generic terms like *instance space* and *concepts*. Broadly speaking an ***instance space*** is typically just the domain of X (your sample population) and an instance of that space is a particular x with various features. For example, X may be the set of all people and an instance of x will be an individual person who is 30 female and medium heights etc.

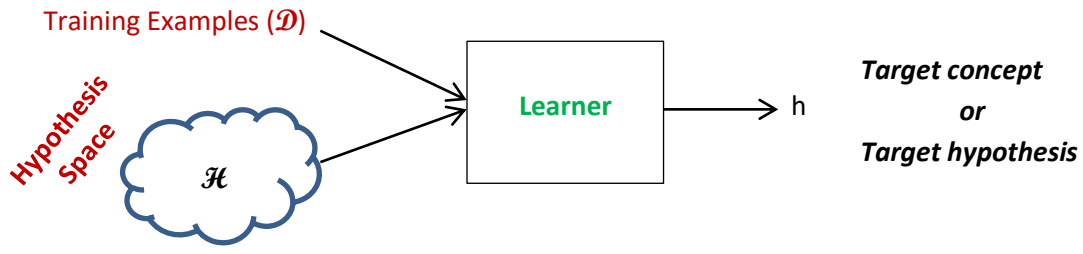
Concept Space

The ***concept space*** on the other hand is subset of the domain X such that X -> {0, 1} or in other words a subset that maps X via Boolean function as either 0 or 1. For example, Our concept space can be people who walk to work. Every person X can either say yes they walk to work (and thus x -> 1) or no (x -> 0).

Hypothesis Space

Hypothesis space is the subset of concept space.

Machine Learning Algorithm learns generalized concept from specific/training examples.



Example:

Let there are two attributes (**temperature** and **blood pressure**) to set the goal that whether the person is sick or not.

Where,
Temperature has three different values L (Low), N (Normal) & H (High)
Blood Pressure has three different values L (Low), N (Normal) & H (High)
Sick has two values Yes (1) & No (0).

Table

Instance Space (X) and Specific instance (x_i) $i.e. x_i \in X$

Temperature (L, N, H)	**Blood Pressure** (L, N, H)	**Sick** (Y=1, N=0)
L	L	0
L	N	0
L	H	1
N	L	1
N	N	0
N	H	1
H	L	0
H	N	1
H	H	1

Concept (c_i)

Sample or Data Space or Training Data $(\mathcal{D})$
& Individual item (d_i) $i.e., d_i \in \mathcal{D}$

Instance Space (X) =|T| x |BP|
Concept Space (C) is of $2^{|T| \times |BP|} = 2^{3 \times 3} = 2^9$ and they are as below:

Hypothesis Space $(\mathcal{H})$

$$\begin{bmatrix}0&0&0\\0&0&0\\0&0&0\end{bmatrix}\begin{bmatrix}1&0&0\\0&0&0\\0&0&0\end{bmatrix}\begin{bmatrix}0&1&0\\0&0&0\\0&0&0\end{bmatrix}\begin{bmatrix}0&0&1\\0&0&0\\0&0&0\end{bmatrix} \dots\dots\dots\dots\dots \begin{bmatrix}1&1&1\\1&1&1\\1&1&1\end{bmatrix}$$

Concept Space (C)

Single item of Concept Space is denoted by c_i and $c_i \in C$

Hypothesis Space $(\mathcal{H})$ is the subset of *Concept Space (C)* i.e. $\mathcal{H} \subset C$ and our model ($h \in \mathcal{H}$).

In above example, |T| = 3 & |BP| = 3

Inxtance Space (X) = |T| x |BP| = 3 x 3 = 9

Data Space ($\mathcal{D}$) =

Concept Space (C) = 2^9 = 512

Hypothesis Space ($\mathcal{H}$)
- Syntactically Distinct Hypothesis Space
- Semantically Distinct Hypothesis Space

Syntactically Distinct Hypothesis Space ($\mathcal{H}$) = (|T| + 2) x (|BP| + 2) = 5 x 5 =25

Semantically Distinct Hypothesis Space ($\mathcal{H}$) = 1 + (|T| + 1) x (|BP| + 1) = 1 + (3 + 1) x (3 + 1) = 1 + 4 x 4 = 1 + 16 = 17