What exactly is a hypothesis space in machine learning?

Asked 5 years, 4 months ago Active 10 months ago Viewed 33k times

Ask Question

Whilst I understand the term conceptually, I'm struggling to understand it operationally. Could anyone help me out by providing an example?

machine-learning terminology definition

Share Cite Edit Follow Flag

edited Jan 2 '19 at 18:29 — user82135

asked Nov 28 '15 at 12:04 — Five ♂ — 365 1 2 11

A space where we can predict output by a set of some legal hypothesis (or function) and function is represented in terms of features. – Abhishek Kumar Aug 9 '18 at 17:05

Add a comment

Start a bounty

3 Answers

Active | Oldest | Votes

Lets say you have an unknown target function $f : X \rightarrow Y$ that you are trying to capture by learning. In order to capture the target function you have to come up with some hypotheses $h_1, \ldots, h_n$ where $h \in H$. Here, H is your hypothesis space or set.

For more information browse Abu-Mostafa's presentaton slides:
https://work.caltech.edu/textbook.html

Share Cite Edit Follow Flag

answered Nov 28 '15 at 13:20 — JimBoy — 920 7 14

7 This answer conveys absolutely no information! What is the intended relationship between f, h, and H? What is meant by "hypothesis set"? – whuber ♦ Nov 28 '15 at 20:50

4 Please take a few minutes with our help center to learn about this site and its standards, JimBoy. – whuber ♦ Nov 28 '15 at 20:57

The answer says very clear. `h` learns to capture target function `f`. `H` is the space where h1, h2,..hn got defined. – Logan Nov 29 '18 at 21:47

Add a comment

Suppose an example with four binary features and one binary output variable. Below is a set of observations:

```
x1 x2 x3 x4 | y
------------|--
 0  0  0  1 | 0
 0  1  0  1 | 0
 1  1  0  0 | 1
 0  0  1  0 | 1
```

This set of observations can be used by a **machine learning (ML) algorithm** to learn a function `f` that is able to predict a value `y` for any input from the **input space**.

We are searching for the ground truth `f(x) = y` that explains the relation between `x` and `y` for all possible inputs in the correct way.

The function `f` has to be chosen from the **hypothesis space**.

To get a better idea: The input space is in the above given example 2^4, its the number of possible inputs. The hypothesis space is $2^{2^4} = 65536$ because for each set of features of the input space two outcomes (`0` and `1`) are possible.

The ML algorithm helps us to find **one function**, sometimes also referred as hypothesis, from the relatively large hypothesis space.

References

- A Few Useful Things to Know About ML

Share Cite Edit Follow Flag

edited Feb 5 '19 at 11:30 — Lerner Zhang — 3,816 1 22 42

answered Sep 24 '17 at 11:57 — So S — 403 4 8

1 Just a small note on your answer: the size of the hypothesis space is indeed 65,536, but the a more easily explained expression for it would be $2^{(2^4)}$, since, there are 2^4 possible unique samples, and thus $2^{(2^4)}$ possible label assignments for the entire input space. – engelen Jan 10 '18 at 9:52

1 @engelen Thanks for your advice, I've edited the answer. – So S Jan 10 '18 at 21:00

@SoS That **one function** is called classifier?? – user125163 Aug 22 '18 at 16:26

2 @Arjun Hedge: Not the one, but one function that you learned is the classifier. The classifier could be (and that's your aim) the one function. – So S Aug 22 '18 at 16:50

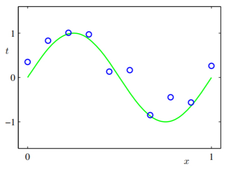
Add a comment

The hypothesis space is very relevant to the topic of the so-called Bias-Variance Tradeoff in maximum likelihood. That's if the number of parameters in the model(hypothesis function) is too small for the model to fit the data(indicating underfitting and that the hypothesis space is too limited), the bias is high; while if the model you choose contains too many parameters than needed to fit the data the variance is high(indicating overfitting and that the hypothesis space is too expressive).

As stated in So S' answer, if the parameters are discrete we can easily and concretely calculate how many possibilities are in the hypothesis space(or how large it is), but normally under realy life circumstances the parameters are continuous. Therefore generally the hypothesis space is uncountable.

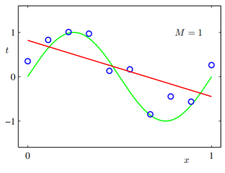
Here is an example I borrowed and modified from the related part in the classical machine learning textbook: Pattern Recognition And Machine Learning to fit this question:

We are selecting a hypothesis function for an unknown function hidding in the training data given by a third person named CoolGuy living in an extragalactic planet. Let's say CoolGuy knows what the function is, because the data cases are provided by him and he just generated the data using the function. Let's call it(we only have the limited data and CoolGuy has both the unlimited data and the function generating them) the ground truth function and denote it by $y(x, w)$.



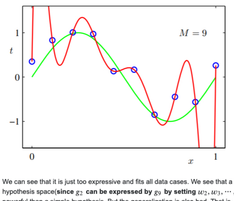
The green curve is the $y(x, w)$, and the little blue circles are the cases we have(they are not actually the true data cases transmitted by CoolGuy because of the it would be contaminated by some transmission noise, for example by macula or other things).

We thought that that hidden function would be very simple then we make an attempt at a linear model(make a hypothesis with a very limited space): $g_1(x, w) = w_0 + w_1 x$ with only two parameters: w_0 and w_1 , and we train the model use our data and we obtain this:



We can see that no matter how many data we use to fit the hypothesis it just doesn't work because it is not expressive enough.

So we try a much more expressive hypothesis: $g_9 = \sum_i^9 w_i x^i$ with ten adaptive paramaters $w_0, w_1 \cdots, w_9$, and we also train the model and then we get:



We can see that it is just too expressive and fits all data cases. We see that a much larger hypothesis space(**since g_1 can be expressed by g_9 by setting $w_2, w_3, \cdots, w_9$ as all 0**) is more powerful than a simple hypothesis. But the generalization is also bad. That is, if we recieve more data from CoolGuy and to do reference, the trained model most likely fails in those unseen cases.

Then how large the hypothesis space is large enough for the training dataset? We can find an aswer from the textbook aforementioned:

> One rough heuristic that is sometimes advocated is that the number of data points should be no less than some multiple (say 5 or 10) of the number of adaptive parameters in the model.

And you'll see from the textbook that if we try to use 4 parameters, $g_1 = w_0 + w_1 x + w_2 x^2 + w_3 x^3$, the trained function is expressive enough for the underlying function $y = \sin(2\pi x)$. It's kind a black art to find the number 3(the appropriate hypothesis space) in this case.

Then we can roughly say that the hypothesis space is the measure of how expressive you model is to fit the training data. The hypothesis that is expressive enough for the training data is the good hypothesis with an expressive hypothesis space. To test whether the hypothesis is good or bad we do the cross validation to see if it performs well in the validation data-set. If it is neither underfitting(too limited) nor overfitting(too expressive) the space is enough(according to Occam Razor a simpler one is preferable, but I digress).

Share Cite Edit Follow Flag

edited May 16 '20 at 15:50

answered Sep 24 '16 at 6:23 — Lerner Zhang — 3,816 1 22 42

This approach looks relevant, but your explanation does not agree with that on p. 5 of your first reference: "A function $h : X \rightarrow \{0, 1\}$ is called [an] hypothesis. A set H of hypotheses among which the approximation function y is searched is called [the] hypothesis space." I would agree the slide is confusing, because its explanation implicitly requires that $C = \{0, 1\}$, whereas that is generically labeled "classes" in the diagram. But let's not pass along that confusion: let's rectify it.) – whuber ♦ Sep 24 '16 at 15:33

1 @whuber I updated my answer just now more than two years later after I have learned more knowledge on the topic. Please help check if I can rectify it in a better way. Thanks. – Lerner Zhang Feb 5 '19 at 11:41

Add a comment

Your Answer

Community wiki

Post Your Answer

Not the answer you're looking for? Browse other questions tagged machine-learning terminology definition or ask your own question.