Neural Networks: What does “linearly separable” mean?

I am currently reading the Machine Learning book by Tom Mitchell. When talking about neural networks, Mitchell states:

> "Although the perceptron rule finds a successful weight vector when the training examples are linearly separable, it can fail to converge if the examples are not linearly separable. "

I am having problems understanding what he means with "linearly separable"? Wikipedia tells me that "two sets of points in a two-dimensional space are linearly separable if they can be completely separated by a single line."

But how does this apply to the training set for neural networks? How can inputs (or action units) be linearly separable or not?

I'm not the best at geometry and maths - could anybody explain it to me as though I were 5? ;) Thanks!

machine-learning neural-network

asked Dec 20 '12 at 16:34
JB2
457 ● 4 ● 13

Regardless of my answer below, I'm pretty sure that book defines "linearly separable" somewhere. Have you tried looking at the Glossary to find where that term is defined? – HerrKaputt Dec 20 '12 at 17:03

@HerrKaputt the definition probably is too mathematical for him... – Anony-Mousse Dec 20 '12 at 17:06

@HerrKaputt I couldn't quite understand the definition given in the book. Thank you for taking the time to explain it to me! – JB2 Dec 20 '12 at 20:16

3 Answers

Suppose you want to write an algorithm that decides, based on two parameters, size and price, if an house will sell in the same year it was put on sale or not. So you have 2 inputs, size and price, and one output, will sell or will not sell. Now, when you receive your training sets, it could happen that the output is not accumulated to make our prediction easy (Can you tell me, based on the first graph if `X` will be an N or S? How about the second graph):

```
        ^
        |  N S   N
       s|  S X    N
       i|  N     N S
       z|  S  N  S  N
       e|  N S  S N
        +----------->
          price


        ^
        |  S S   N
       s|  X S    N
       i|  S     N N
       z|  S  N  N  N
       e|    N N N
        +----------->
          price
```

Where:

```
S-sold,
N-not sold
```

As you can see in the first graph, you can't really separate the two possible outputs (sold/not sold) by a straight line, no matter how you try there will always be both `S` and `N` on the both sides of the line, which means that your algorithm will have a lot of `possible` lines but no ultimate, correct line to split the 2 outputs (and of course to predict new ones, which is the goal from the very beginning). That's why `linearly separable` (the second graph) data sets are much easier to predict.

edited Jul 27 at 16:41
Duck
24 ● 6

answered Dec 20 '12 at 17:49
Valentin Radu
5,003 ● 7 ● 31 ● 75

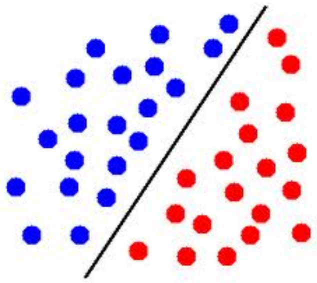
Another great explanation. Thank you very much Valentin! – JB2 Dec 20 '12 at 19:45

How would this apply to a continuous result and not a classification result ? For example if my inputs were distance and slope and my output was fuel consumed? – user3061923 Feb 14 '14 at 8:00

This means that there is a hyperplane (which splits your input space into two half-spaces) such that all points of the first class are in one half-space and those of the second class are in the other half-space.

In two dimensions, that means that there is a line which separates points of one class from points of the other class.

EDIT: for example, in this image, if blue circles represent points from one class and red circles represent points from the other class, then these points are linearly separable.



In three dimensions, it means that there is a plane which separates points of one class from points of the other class.

In higher dimensions, it's similar: there must exist a hyperplane which separates the two sets of points.

You mention that you're not good at math, so I'm not writing the formal definition, but let me know (in the comments) if that would help.

edited Dec 20 '12 at 17:11

answered Dec 20 '12 at 16:40
HerrKaputt
1,409 ● 5 ● 15

This is a great explanation. Thank you! – JB2 Dec 20 '12 at 19:44

Look at the following two data sets:

```
^                           ^
|   X     O                 |  AA     /
|                           |  A     /
|                           |       /   B
|   O     X                 |  A   /    BB
|                           |     /   B
+----------->               +----------->
```

The left data set is *not* linearly separable (without using a kernel). The right one is separable into two parts for `A' and` B` by the indicated line.

I.e. You cannot **draw a *straight* line** into the left image, so that all the `X` are on one side, and all the `O` are on the other. That is why it is called "not linearly separable" == there exist no linear manifold separating the two classes.

Now the famous kernel trick (which will certainly be discussed in the book next) actually allows many linear methods to be used for non-linear problems by virtually adding additional dimensions to make a non-linear problem linearly separable.

edited Dec 20 '12 at 17:06

answered Dec 20 '12 at 16:42
Anony-Mousse
31.2k ● 3 ● 44 ● 88