INDUCTIVE BIAS(RESTRICTIONS/CONTRAINTS/PRIOR ASSUMPTIONS/PREFERENCES)

Consider, for example, the instances X and hypotheses H in the EnjoySport learning task. Given that the attribute **Sky** has three possible values, and that **AirTemp, Humidity, Wind, Water**, and **Forecast** each have two possible values, the instance space X contains exactly 3 2 2 .2 2 .2 = 96 distinct instances.

A similar calculation shows that there are 5.4.4.4.4.4 = 5,120 syntactically distinct hypotheses within H. Notice, however, that every hypothesis containing one or more "IZI" symbols represents the empty set of instances; that is, it classifies every instance as negative. Therefore, the number of semantically distinct hypotheses is only 1 + (4.3.3.3.3.3) = 973.

**H**  =

Target concept is in the hypothesis space H is to provide a hypothesis space capable of representing every teachable concept ( C(X) H)

X of days described by the six available attributes is 96.

How many possible concepts can be defined over this set of instances?

In other words, how large is the power set of X?

In general, the number of distinct subsets that can be defined over a set X containing |x| elements (i.e., the size of the power set of X) is 2|x|.

Thus, there are 2**96**, or approximately~ 1028 Target Concept

**UNBIASED LEARNER**

****

****

This hypothesis, although it is the maximally specific hypothesis from H that is consistent with the first two examples, is already overly general: it incorrectly covers the third (negative) training example. The problem is that we have biased the learner to consider only conjunctive hypotheses. In this case we require a more expressive hypothesis space.

****

To see why, suppose we present three positive examples (x1,x2,x3) and two negative examples (x4, x5) to the learner. At this point, the S boundary of the version space will contain the hypothesis which is just the disjunction of the positive examples. The G boundary will consist of the hypothesis that rules out only the observed negative examples

****

****

Therefore, the only examples that will be unambiguously classified by ***S*** and G are the observed training examples themselves.

**FUTILITY OF BIAS FREE LEARNER**

***A learner that makes no a priori assumptions regarding the identity of the target concept has no rational basis for classifying any unseen instances***.

In fact, the only reason that the CANDIDATE-ELIMINATION algorithm was able to generalize beyond the observed training examples in our original formulation of the EnjoySport task is that it was biased by the implicit assumption that the target concept could be represented by a conjunction of attribute values.

In cases where this assumption is correct (and the training examples are error-free), its classification of new instances will also be correct. If this assumption is incorrect, however, it is certain that the CANDIDATE-ELIMINATION algorithm will misclassify at least some instances from X.

The key idea we wish to capture here is the policy by which the learner generalizes beyond the observed training data, to infer the classification of new instances.

Therefore, consider the general setting in which an arbitrary learning algorithm L is provided an arbitrary set of training data **D = {(x, c(x))}** of some arbitrary target concept **c**. After training, L is asked to classify a new instance Xi.

Let L(xi, D,) denote the classification (e.g., positive or negative) that L assigns to xi after learning from the training data D,. We can describe this inductive inference step performed by L as follows



Where the notation y > z indicates that z is inductively inferred from y.

For example, if we take L to be the CANDIDATE-ELIMINATION algorithm, D, to be the training data from the following table, and xi to be the first instance from the following example instances, then the inductive inference performed in this case concludes that L(xi, D,) = (EnjoySport = yes).



More precisely, we define the inductive bias of ***L*** to be the set of assumptions ***B*** such that for all new instances ***xi*.**



Where the notation y t z indicates that z follows deductively from y (i.e., that z is provable from y). Thus, we define the inductive bias of a learner as the set of additional assumptions B sufficient to justify its inductive inferences as deductive inferences.

Given this definition of L(xi, D,)

for the CANDIDATE-ELIMINATION Algorithm, what is its inductive bias? It is simply

the assumption c E H. Given this assumption, each inductive inference performed

by the CANDIDATE-ELIMINATION Algorithm can be justified deductively.

To see why the classification L(xi, D,) follows deductively from B = {c E H), together with the data D, and description of the instance xi, consider the following argument.

First, notice that if we assume c E H then it follows deductively that c E VSH,Dc.

This follows from c E H, from the definition of the version space VSH,Dc as the set of all hypotheses in H that are consistent with Dc and from our definition of Dc = {(x, c(x))} as training data consistent with the target concept c.

Second, recall that we defined the classification L(xi, D,) to be the unanimous vote of all hypotheses in the version space. Thus, if L outputs the classification L(xi D,), it must be the case the every hypothesis in VSH,Dc also produces this classification, including the hypothesis c E VSH,Dc.

Therefore c(xi) = L(xi, D,). To summarize, the CANDIDATE-ELIMINATION algorithm defined in this fashion can be characterized by the following bias.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| X1 | Japan | Honda | Blue | 1980 | Economy | Yes |
| X2 | Japan | Toyota | Green | 1970 | Sports | No |
| X3 | Japan | Toyota | Blue | 1990 | Economy | Yes |
| X4 | USA | Chrysler | Red | 1980 | Economy | No |
| X5 | Japan | Honda | White | 1980 | Economy | Yes |

STEP-0 : S:< (‘0’, ‘0’, ‘0’, ‘0’, ‘0’) >

G:< ( ‘?’, ‘?’, ‘?’, ‘?’, ‘?’) >

STEP-1 : S:<( Japan, Honda, Blue, 1980, Economy)>

G:< ( ‘?’, ‘?’, ‘?’, ‘?’, ‘?’) >

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Japan | Toyota | Green | 1970 | Sports | No |

INSTANCE X2:

STEP-2: S:<( Japan, Honda, Blue, 1980, Economy)>

G:< (?, Honda, ?, ?,?), (?, ?, Blue, ?,?), (?, ?, ?, 1980,?), (?, ?, ?, ?, Economy)>

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Japan | Toyota | Blue | 1990 | Economy | Yes |

INSTANCE X3:

STEP-3:

S:<( Japan, ?, Blue,?, Economy)>

G:< (?, ?, Blue, ?,?), (?, ?, ?, ?, Economy)>

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| USA | Chrysler | Red | 1980 | Economy | No |

INSTANCE X4:

STEP-4: S:<( Japan, ?, Blue,?, Economy)>

G:< (Japan, ?, Blue, ?,?), (Japan, ?, ?, ?, Economy), (?, ?, Blue, ?, Economy)>

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Japan | Honda | White | 1980 | Economy | Yes |

INSTANCE X5 :

STEP-5:

S: < (Japan,?, ?, ?, Economy)

G:< (Japan, ?, ?, ?, Economy)