Version Space Algebras and Category Theory

Bailey Wickham

February 3, 2022

Abstract

We give a category theoretic definition of version space algebras and their operations.

1 Introduction

Version Space Algebras (VSAs) are useful as a framework for machine learning in Computer Science, however their associated definitions are limited by this applied context. Many of the constructions can be generalized to be category theoretic.

2 Version Spaces

[] defines a Version Space $VS_{H,D}$ as such:

Definition 2.1 (Version Space)**.** Let the hypothesis space $H = \{f | f : A \to B\}$, where A and B are sets. Let the examples $D = \{(i, o) | i \in A \text{ and } o \in B\}$ Then our version space $VS_{H,D} = \{f | f(i) = o \text{ for all } (i, o) \in D\}$

A version space algebra provides us with a type of "ambient space", H. H provides a context for which the functions exists. The examples, D, are pairs of elements in the domain and range of these functions. Our version space consists of the functions in the ambient space that are consistent with the examples.

In practice the ambient space may be something like "words in our document", or "integers representing row and column numbers", however then only need be functions between (small) sets. Version spaces are used to define all possible actions of a specific type, then constrained to the examples observed in the current document. This is used to generate predictions, where the examples we have used in our document are used to predict future actions.

[] introduces the idea of Version Space Algebras (VSAs) which are version spaces generated by operations on other version spaces. A version space which is defined explicitly, or not defined in terms of operations on other version spaces, is called an *atomic version space*.