

We can convert categorical variables into numeric form with one of two methods:

1. Label encoding: here we simply convert each category into an integer
2. One hot encoding: here we look at the number of total categories possible and convert them into vectors.

Suppose we have

+1 A

+1 A

+1 A

-1 B

-1 B

-1 A

Label encoding:  $A \rightarrow 1, B \rightarrow 2$

One hot encoding  $A \rightarrow (1,0), B \rightarrow (0, 1)$

Thus while one hot encoding increases dimensionality it also gives better signal and likely to be more accurate.