

Lecture 12:

Think about how range-type variables might be represented. Is it reasonable to represent them as any numeric variables or is it better to have ad-hoc encoding?

Representing range-type variables as Numeric variables might be problematic, when we want to compare between different features with different ranges.

For example, one feature has a range of 0-10 and value 8, and a second feature with range of 0-100 and value 8. Some Models might mistake the values to have the same "affect", even so the first is 80% of the range, and the second only 8%.

For this reason we might consider "scaling" the value for the specific range. For instance,

if we have $[a, b]$ range we can consider: $h(x) = \frac{x-a}{b-a}$

This helps us understand the relative size of the value, with the perspective of the range.