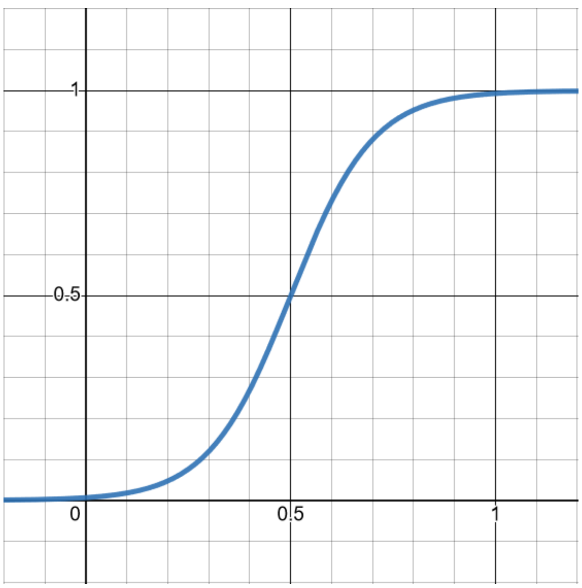
CS x260 Program 3 artificial data explanation (Kyle)

The data provided is a mapping of counts of themes, presumably obtained from generated roadtrips, to a real-valued utility value that corresponds to the user's relative preference for that roadtrip. The utility function used will be referred to as $S(T)$. Theme counts are represented as a list of integer values $T = \{t_1, t_2, \ldots t_n\}$. Users will often have preferences that are complex and interacting in nature. In other words, the utility of combinations of themes may have higher preference than the sum of those themes in isolation; i.e. $S(\{t_i, t_j\}) \geq S(\{t_i\}) + S(\{t_j\})$. Some combinations will also be preferred over others by a given user.

From this, I decided to use the formula to define $S(T) = (1 + R) \sum_{i \subset T} w_i \, f(\omega_i)$. In this formula, i is a subset of the themes in T, w_i is a weight that describes the user's relative preference for that theme subset s.t. $\sum_i w_i = 1$, and $f(\omega_i)$ is a function that maps theme counts for i to an individual preference value, and R is a uniform random variable that injects some random noise. f is shared between all users and all theme subsets, so learning of user preferences amounts to learning the weights w_i.



The function behind $f(\omega_i)$ is a standard logistic function defined as $\frac{1}{1+e^{-c(x(\omega_i)-a)}}$ where $x(\omega_i)$ is a function that squashes theme counts into a scalar value that will be defined below, $c = 10$ controls the slope, and $a = \frac{1}{2}$ is the center value of the function. The graph of this function is shown above. A logistic function was chosen for two primary reasons. First, logistic functions are non-linear and a weighted sum of logistic functions thus cannot be learned by a single layer neural network. Second, the overall utility formula becomes similar, but not identical, to the expected human preference functions defined by Tversky and Kahneman in economic Prospect Theory.

The final piece in need of definition is $x(\omega_i)$. This function is the squashed input to the preference function for a single theme relationship. This is effectively how strongly the relationship between those themes shows in the datapoint. Intuitively, a relationship between t_i and t_j is strongest when both are present and they are similar in number. From this, I used the function $x(\omega_i) = \frac{\min(j, t_k)}{\max(t_j, t_k)} (x(\omega_i) = 0$

when both t_j and t_k are 0 to handle division by 0). This function is highest when $t_i = t_j$ and trend towards 0 as the values diverge. This is desirable because a relationship where one value is high while the other low is one in which that relationship is not strongly represented. A limitation to this approach is that there is no preference for the size of a relationship's presence. That is to say, given the values $\{t_j = 1, t_k = 4, t_m = 3, t_n = 1\}$, the relationship $\{t_j, t_n\}$ will yield a higher value than $\{t_k, t_m\}$ despite the higher average counts in the latter. This is also only defined for size two relations, but can be generalized to the following form:

$$x(t_1, t_2, \dots, t_n) = \frac{2}{n(n-1)} \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} \frac{\min(t_j, t_k)}{\max(t_j, t_k)}$$

Where $\frac{\min(t_j,t_k)}{\max(t_j,t_k)} = 0$ when $t_j = t_k = 0$ to handle division by zero. Intuitively, this is the average activation of every pair of themes in the relationship. The maximum value will be 1 when all themes are equally represented and approaches 0 as more themes differ from each other by larger amounts or are absent entirely.

The formulae described above are restated below for convenience.

$$T = \{t_1, t_2, \dots t_n\}$$

$$S(T) = (1 + R) \sum_{i \subset T} \frac{w_i}{1 + e^{-10(x(\omega_i) - 0.5)}}$$

$$R \sim U_{[-0.1, 0.1]}$$

$$0 \le w_i \le 1 \qquad \sum_i w_i = 1$$

$$x(t_1, t_2, \dots t_n \in \omega_i) = \frac{2}{n(n-1)} \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} \frac{\min(t_j, t_k)}{\max(t_j, t_k)}$$