

# ML Concept for The Wine-inator project

# Data extraction approach

- We are pursuing different options for data including
  - What-to-Pair - <https://whattopair.com/>
  - Spoonacular api- <https://spoonacular.com/food-api>
- What-to-Pair is currently lowest risk approach
  - Will scrape the web pages for the data
  - If Spoonacular api comes through we can pivot or leverage as another component of the app

# What-to-Pair: Data exploration

- 4 columns, Pairs beverage to food choice by 5-star rating

Beverage Category	Beverage	Food	Rating
Red Wine	Pomerol	Red Miso Okazu	Rating (stars)
White Wine	<a href="#">Chardonnay</a>	<a href="#">Hamburger with Grilled Pineapple</a>	4
Red Wine	<a href="#">Montepulciano</a>	<a href="#">Beef Bolognese</a>	4
Dessert Wine	Recioto di Soave Classico Spumante (DOCG)	<a href="#">Aioli</a>	4
White Wine	Garnacha Blanca	<a href="#">Aioli</a>	4
White Wine	<a href="#">Grüner Veltliner</a>	<a href="#">Aioli</a>	3.5
White Wine	<a href="#">Savennières</a>	<a href="#">Aioli</a>	3.5
White Wine	Verdelho	<a href="#">Aioli</a>	3.5

# Question to be addressed by ML predictor

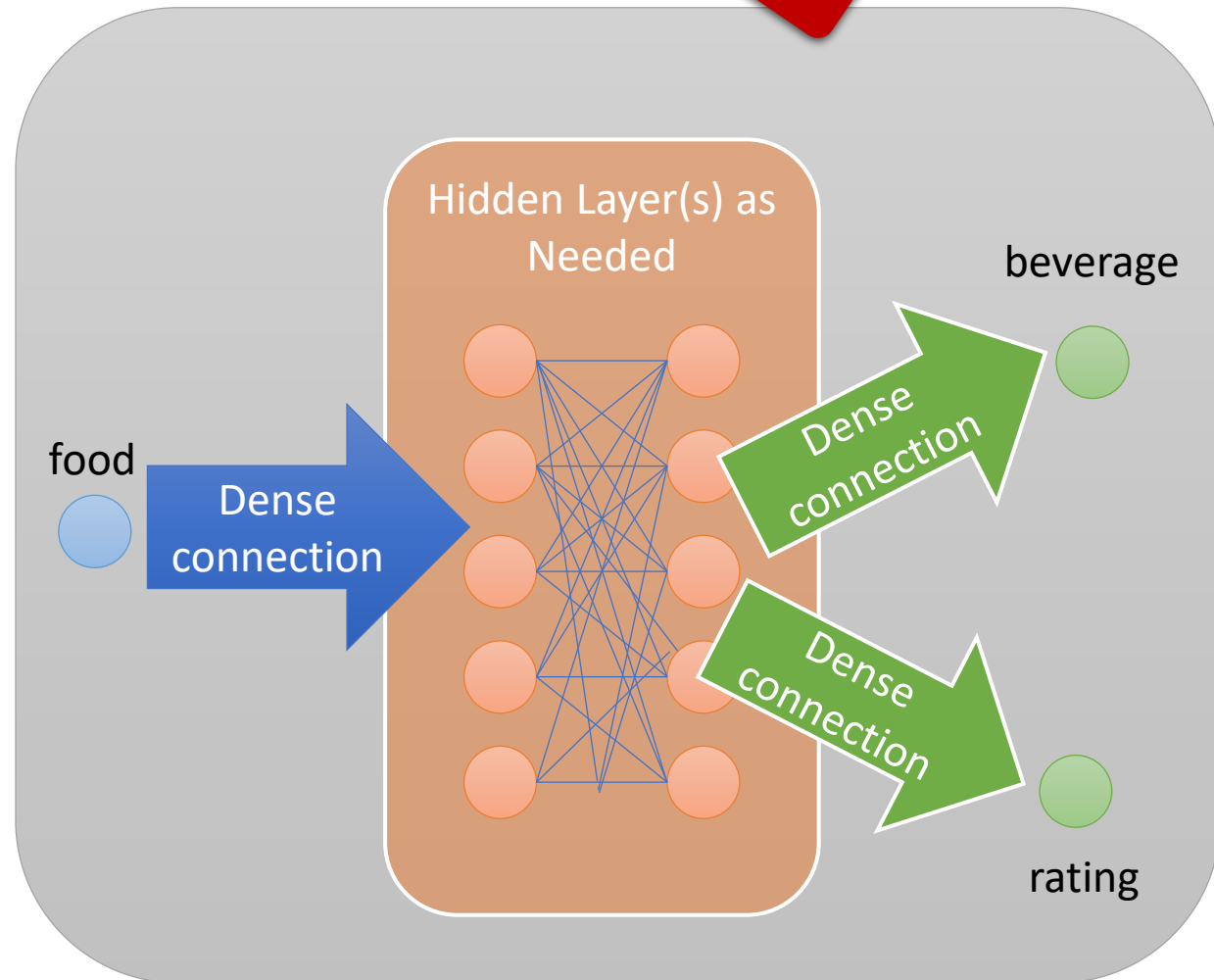
- Assuming What-to-Pair data.
- Question being addressed by ML
  - What wine to pair with a combinations of foods that make up a meal
  - Prediction / Recommendation
    - Input: Select a few foods as input
    - Output: 1 wine or set of wines recommended for the input foods

# ML Concept #1 Will this work? What does it do?

Concept Rejected

## Structure

- 1 Input
    - Food (category encoded integer from 1 to m (where m = index of food))
  - 2 Outputs =
    - Beverage (category encoded integer from 1 to n (where n = index of beverage))
    - Rating (float number from 0 thru 5)
  - Hidden layer (as needed)
- 
- Questions / Notes
    - Will this structure work?
      - E.g do I really know whether if one of the outputs will be a beverage and the other will be a rating?
      - I suspect not.
    - Interesting in terms of information encoding a database
    - Does this do anything more than filtering the database by a food?
    - It does not allow combination of food inputs



# ML Concept #2: Predictor using one hot encoded inputs and outputs

## Structure

- m Inputs
  - A node for every food option
  - Use to-categorical encoded inputs
- n Outputs
  - A node for every beverage option
  - Use to-categorical encoded outputs
- Hidden layer (as needed)
- Not sure what to do with Rating data
- Notes
  - Train the net using food x beverage records
  - Filter data by rating  $\geq 3$  stars
  - Test using individual foods (1 food at a time)
  - Predict by turning on multiple foods as inputs
    - Let the net do the work of combining the data.
    - Sort wine results by level of activation and present hottest activated wine nodes
- Questions
  - Is this concept sound?
  - Is there a way to use the Rating data to weight the training of the net

