

INTRODUCTION

Motivation: Food, healthy eating and new recipes have become central subjects in our daily life. Specially with obesity being a major health problem across the world, specially in the US and an increased public interest in healthy lifestyles, the study of techniques to assist people following a healthy diet is significant.



- What other ingredients can go with blueberry cheesecake?
- Will those ingredients be **healthy**?
- We got a **SOLUTION!**

Goals:

- Recommend top k ingredients that can be added to a partial recipe
- Consider health values (calories, protein, fat)
- Have fun!

DATA PROCESSING

Data Collected: 'Epicurious - Recipes With Nutrition' from Kaggle [1]

- 27k Recipes
- 680 Columns

Data Processing:

- Removed columns that were not ingredients
- Removed rows with missing values

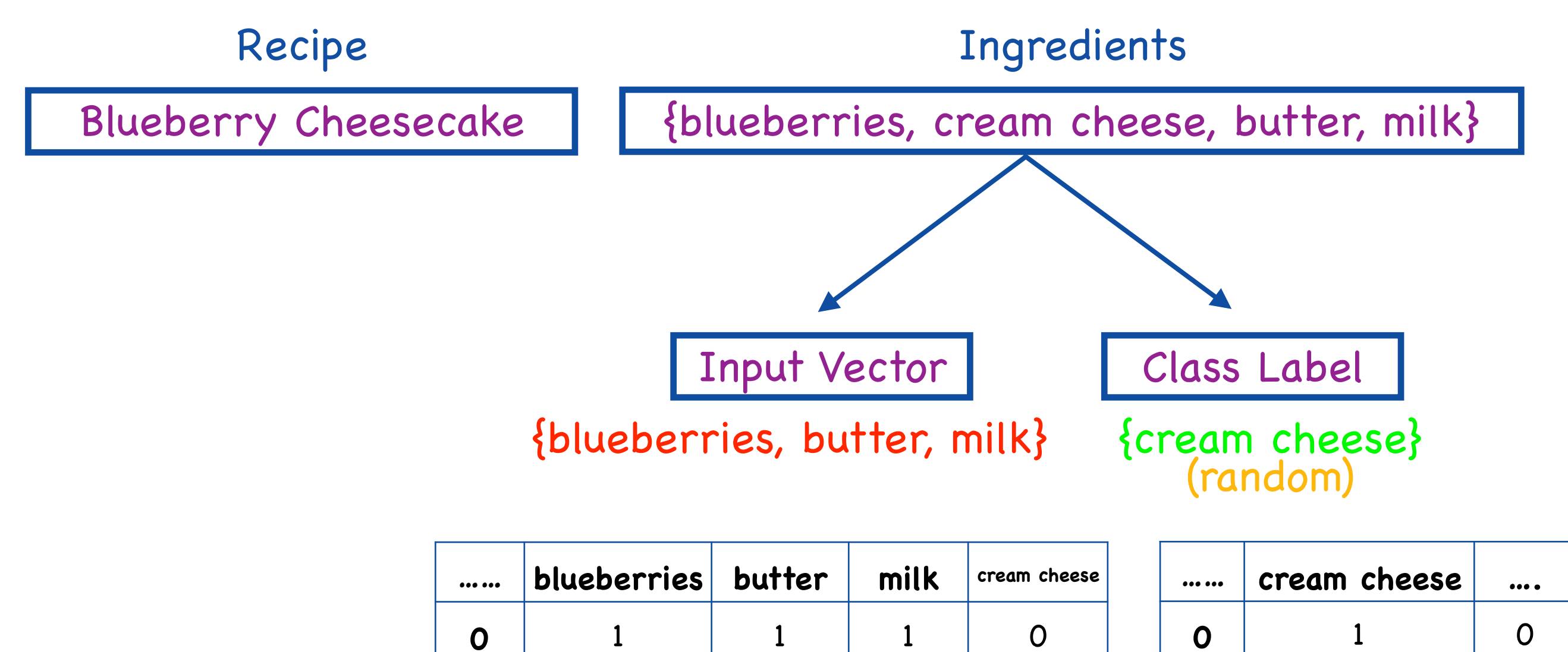
Processed Output:

title	rating	calories	protein	fat	apple	...	banana	...	milk
lentil, apple and turkey wrap	4.5	426.0	30.0	7.0	1		0		0
Potato and fennel hodge	3.75	165.0	6.0	7.0	0		0		0
Blueberry Cheesecake	5.0	180.0	8.0	15.5	0				1
Korean Marinated Beef	4.375	170.0	7.0	10.0	0		0		0

- The columns contain health values and all possible list of ingredients
- The rows are recipes with their health values and ingredients
- If one recipe contains one ingredient, the corresponding value is 1, otherwise 0

NEURAL NETWORKS

We have implemented a simple neural network model with 308 nodes in input and output layer, 300 nodes in hidden layers and total of 4 dense hidden layers.

Training Phase:

- Loss Function: Categorical Cross- Entropy
- Optimizer: adam
- Activation Function: sigmoid
- 80% training set, 20% test set
- 18 epochs

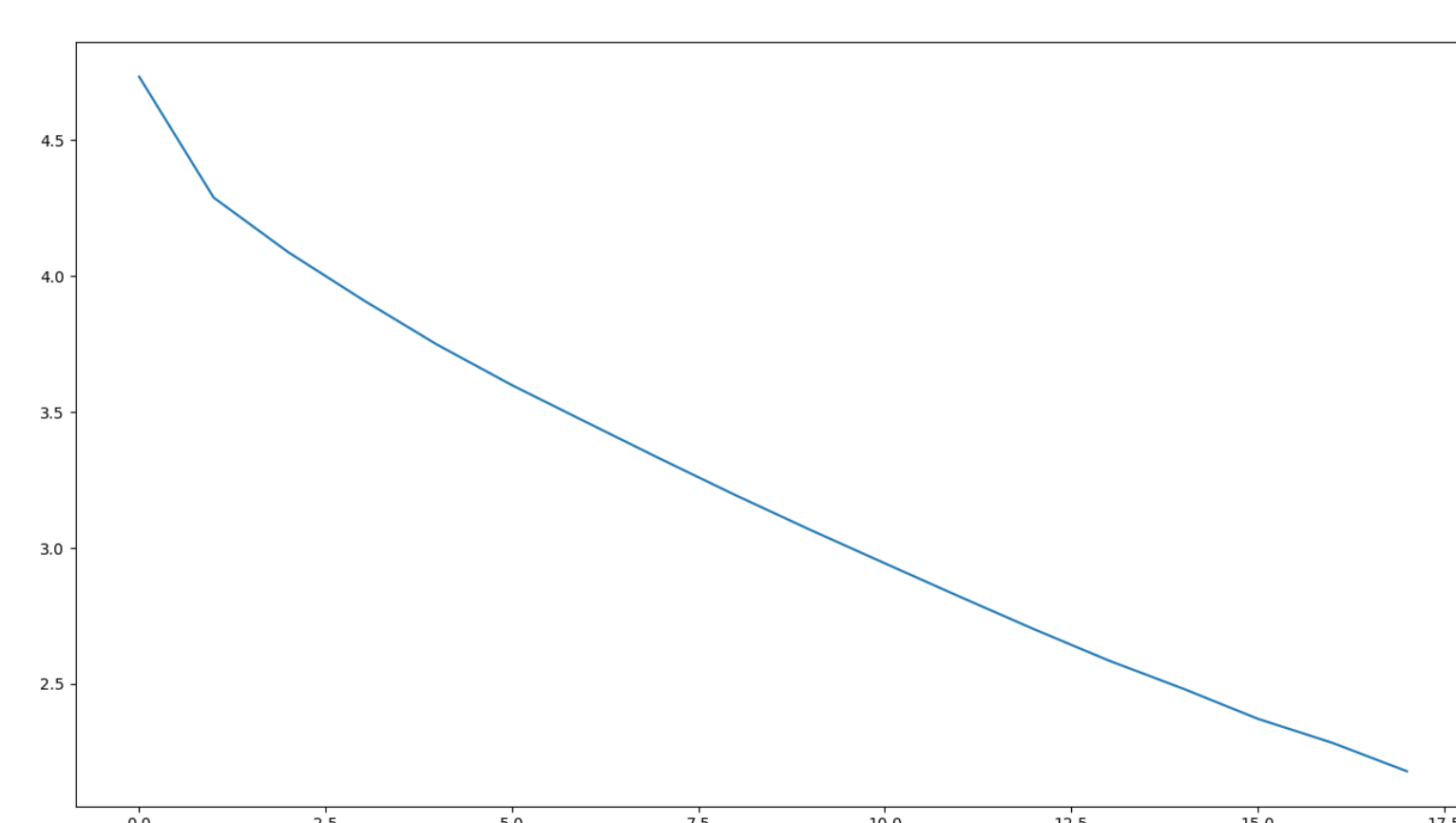
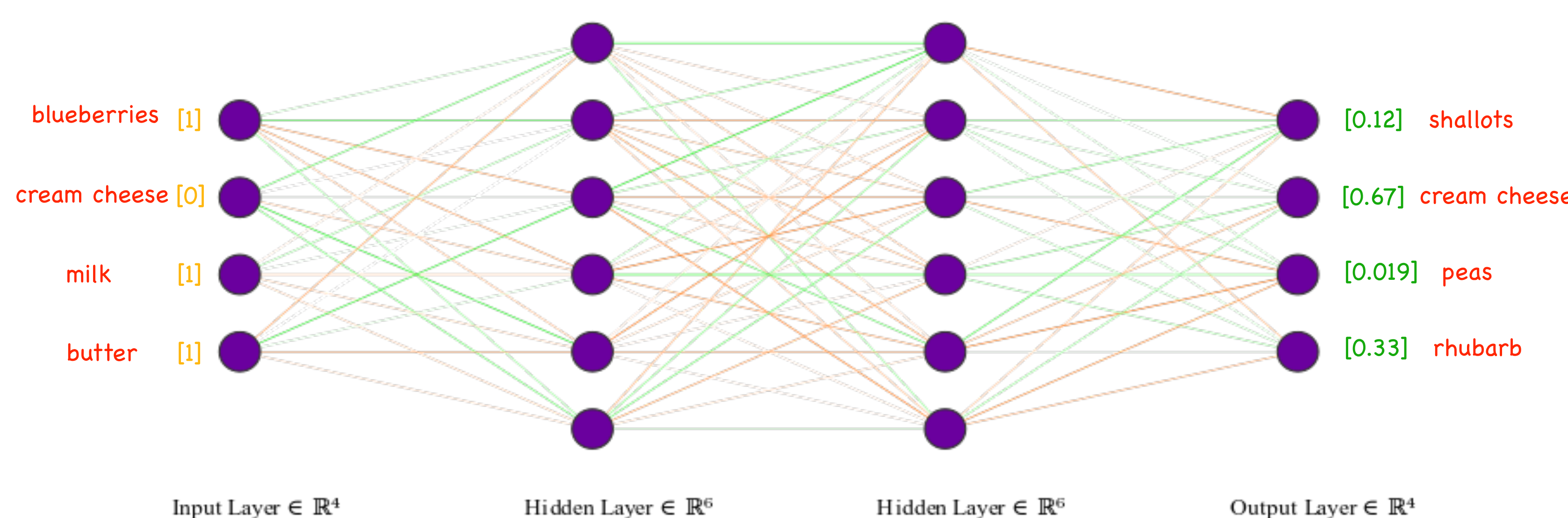


Fig: Loss function reaching minima

Output: For given input of 'blueberry cheesecake', our model suggests the following:

Ingredient	Score
cream cheese	0.76
chocolate	0.64
mint	0.55
oat	0.54

- Suggests 10 ingredients (here 5 is shown)
- Score is confidence (correlation among ingredients)
- **Cream cheese** was originally in the recipe!

APRIORI ALGORITHM

- We also implemented apriori algorithm to see correlations among ingredients and their tags
- Finds correlation among ingredients and their tags which produces interesting results

Support	Confidence	Rule
0.24	0.7273	{'kosher', 'soy free'} -> {'peanut free', 'vegetarian'}
0.25	0.7353	{'vegetarian'} -> {'pescatarian', 'fish'}
0.26	0.7567	{'lentil', 'salad'} -> {'carrots'}
0.26	0.7571	{'blueberry', 'cream cheese'} -> {'dessert', 'cake'}

- Given a min support and min confidence, generates rules $X \rightarrow Y$
- Rules are the correlation among ingredients

EVALUATION

For evaluation we are using leave-one-out cross validation scheme.

- Take one ingredient out of recipe
- If that ingredient comes up in the top 5 suggested ingredients, that's a positive result
- **77% Accuracy!**
- For categorical accuracy (only top ingredient is counted) : **47%**

DISCUSSIONS

- The accuracy is not very high because our dataset contains only 27k recipes. Higher accuracy can be achieved if we have a larger dataset
- Accuracy in terms of "food taste" is subjective, it may happen that ingredients other than the 'top' is tastier to some people

REFERENCES

- [1] <https://www.kaggle.com/hugodarwood/epirecipes>

HOW DOES HEALTH COME IN

But wait! Our project does something novel. After getting the sorted order of suggested ingredients, we pass it through our health function to **re-suggest** according to health value.

- Add a new column in dataset called healthVal

$$healthVal = \frac{protein}{calorie + fat}$$

- Train a neural network again
- Predict healthVal after adding each suggested ingredient
- Re-sort the order according to healthVal

Ingredient	healthVal
mint	0.87
oat	0.85
cream cheese	0.48
chocolate	0.33

