UMSL

Project Report for CS 4300

Contraceptive Methods for Women in Indonesia

Hana A Abuzeid
Prof. Badhikari
Fall 2020

Department of Mathematics and Computer Science

December 16, 2020

Contents

1 Introduction

What are Contraceptives?

Contraceptives are a form of birth control women take to prevent pregnancy. There are many contraceptives for example: pills, patches, injections, IUD's etc...

Why this topic?

I currently work as a pharmacy technician and in our pharmacy most females take a form of contraceptive. And it interests me to see the statistical values and predictions we'll get based on women's demography and social life characteristics. predication's are which is mostly used by women (no use, long-term methods, or short-term methods).

2 Data Set

This data set is a subset of the 1987 National Indonesia Contraceptive Prevalence Survey. These are samples of married women who were either not pregnant or do not know if they were at the time of interview.
It contains 1473 samples in the data set. Each row list is described at are 10. However, after cleaning the set and removing unnecessary factors (such as, husband education, wife working and wife religion) as well as sorting out the 3 inputs which is the contraceptive methods (making them binary, so we remove an input. short use, long use, and no use) calculating which ones are relevant we kept the short use and no use.Now we have 6 features and 2 inputs. The Contraceptive method and other features are labeled below.

Input Features:

1. Wife's age (numerical)
2. Wife's education (categorical) 1=low, 2, 3, 4=high
3. Husband's education (categorical) 1=low, 2, 3, 4=high
4. Number of children ever born (numerical)
5. Wife's religion (binary) 0=Non-Islam, 1=Islam
6. Wife's now working? (binary) 0=Yes, 1=No
7. Husband's occupation (categorical) 1, 2, 3, 4
8. Standard-of-living index (categorical) 1=low, 2, 3, 4=high
9. Media exposure (binary) 0=Good, 1=Not good
10. Contraceptive method used (class attribute) 1=No-use, 2=Long-term, 3=Short-term

UCI Data Set URL: `https://archive.ics.uci.edu/ml/datasets/Contraceptive+Method+Choice`
Google Colab: `https://colab.research.google.com/drive/15e2Uy8RaLp30VZ6uzm72SmmPng1dhS4x?usp=sharing`

2.2 Visualization of each input data? Our Data is a binary data set as we mentioned before we categorized the input as 0- no use and 1- for short term use and the figure below shows that statistical values of each of our input with the mean, st, min, and max.

```
1  data.describe()
```

	wife_age	wife_education	num_of_children	husband_occupation	standard_living	media_exposure	contraceptive_method
count	1473.000000	1473.000000	1473.000000	1473.000000	1473.000000	1473.000000	1473.000000
mean	32.538357	2.958588	3.261371	2.137814	3.133741	0.073999	1.919891
std	8.227245	1.014994	2.358549	0.864857	0.976161	0.261858	0.876376
min	16.000000	1.000000	0.000000	1.000000	1.000000	0.000000	1.000000
25%	26.000000	2.000000	1.000000	1.000000	3.000000	0.000000	1.000000
50%	32.000000	3.000000	3.000000	2.000000	3.000000	0.000000	2.000000
75%	39.000000	4.000000	4.000000	3.000000	4.000000	0.000000	3.000000
max	49.000000	4.000000	16.000000	4.000000	4.000000	1.000000	3.000000

Figure 1: Input Feature Statistics

3 Normalize Data

Normalization is a probability function that depicts how the estimations of a variable are distributed. It is a symmetric distribution where the greater part of the perceptions cluster around the focal area and the probabilities for values separates away from the mean equally in both sides. Figure 2 below is the data being normalized.

$$X = \frac{X - X\,\min}{X\,\max - X\,\min}$$

About our data: Since it is a binary cross entropy/classification we use stigmoids, purpose of it because its a real valued number to a probability between 0 and 1. We then extract data from file and return it into an numpy.ndarray, we prepare the data by splitting it into input and output normalizes input using min-max normalizing. Accuracy in a multi layer is about 86% but with a single layer (logistic regression) its about 92% which shows that its more accurate on a single layer feature.

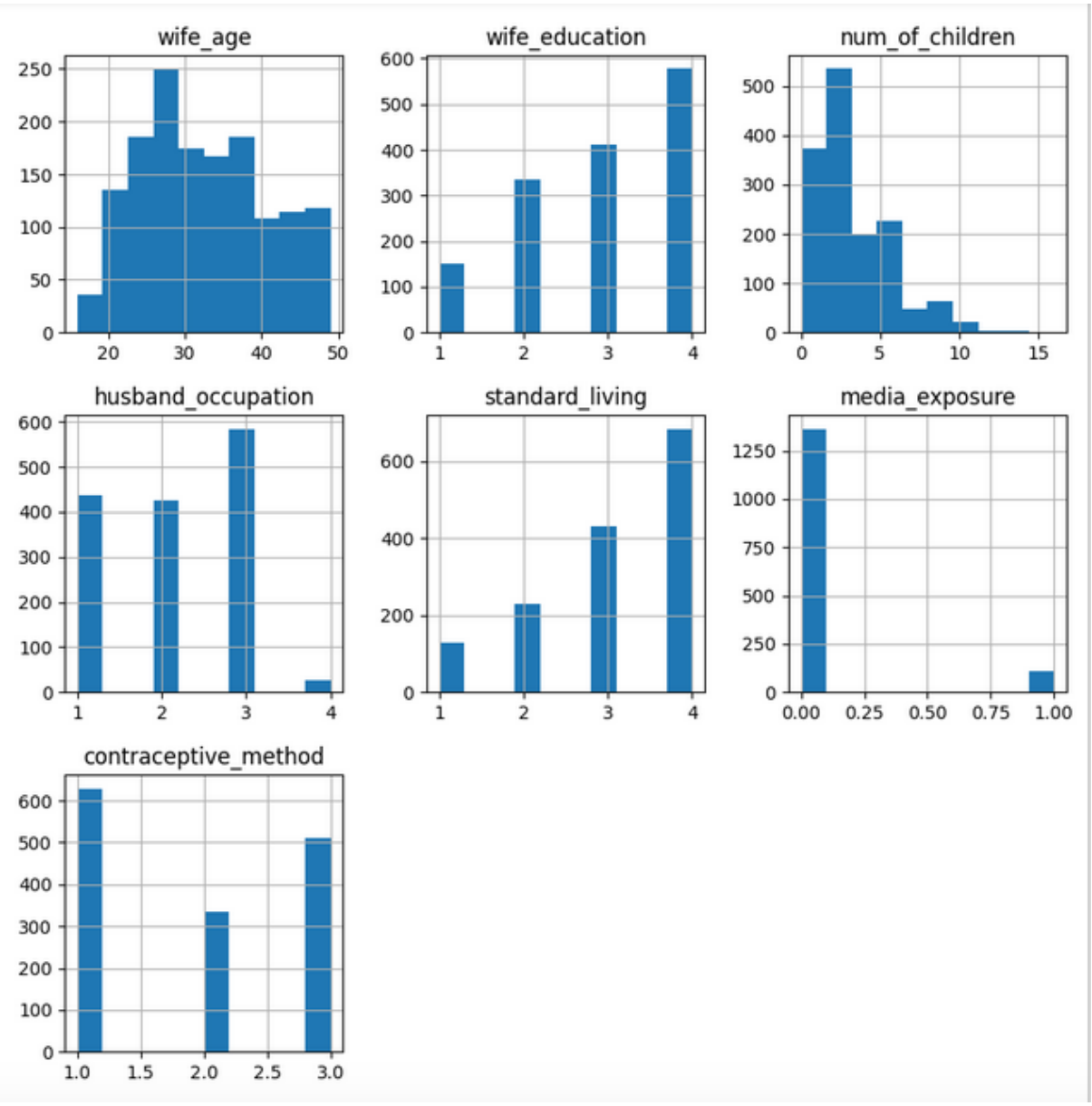

Figure 2: Feature Statistics for different life characteristics

4 Model and Evaluation

Data Splitting

Based on the data that was randomly shuffled, the data set was split to training and validation. 93% was allocated for training and 42% was allocated for validation.

Learning curve

Learning curves shows us the behavior and how fast the machine can improve with the increase of the number of training. Show below is the result and how there is a big gap in between our data.

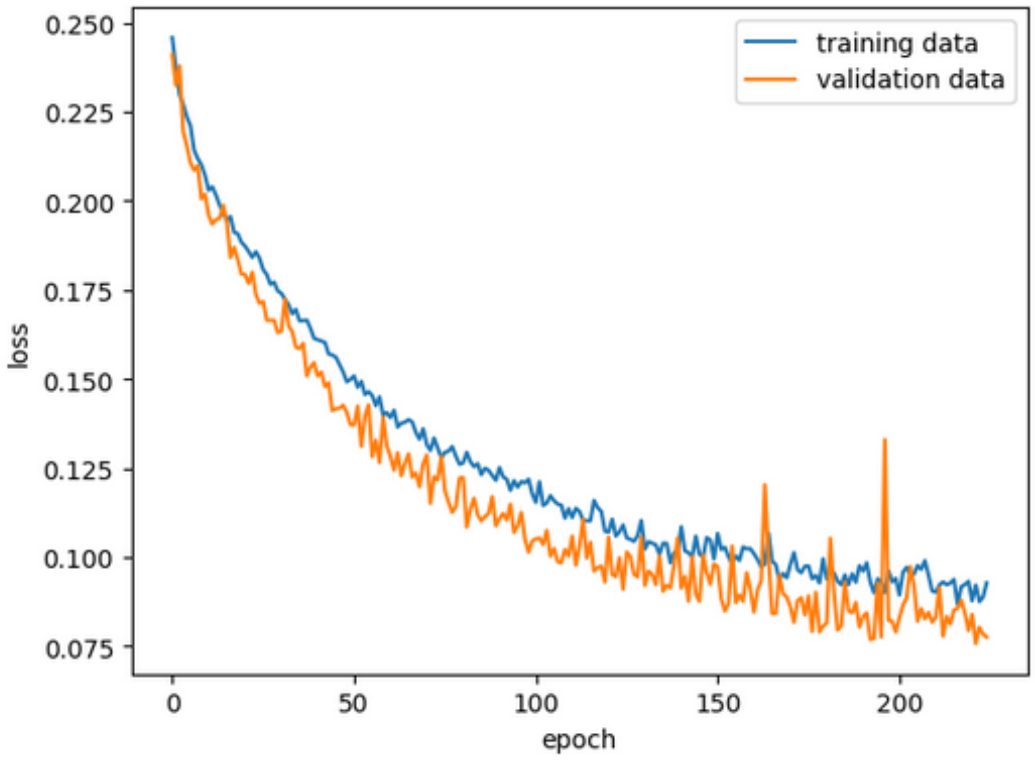

Figure 3: Learning Curve

The best learning curve that I've done during the testing stage was in figure 4 because its smoother than other figures and no over fitting was present.

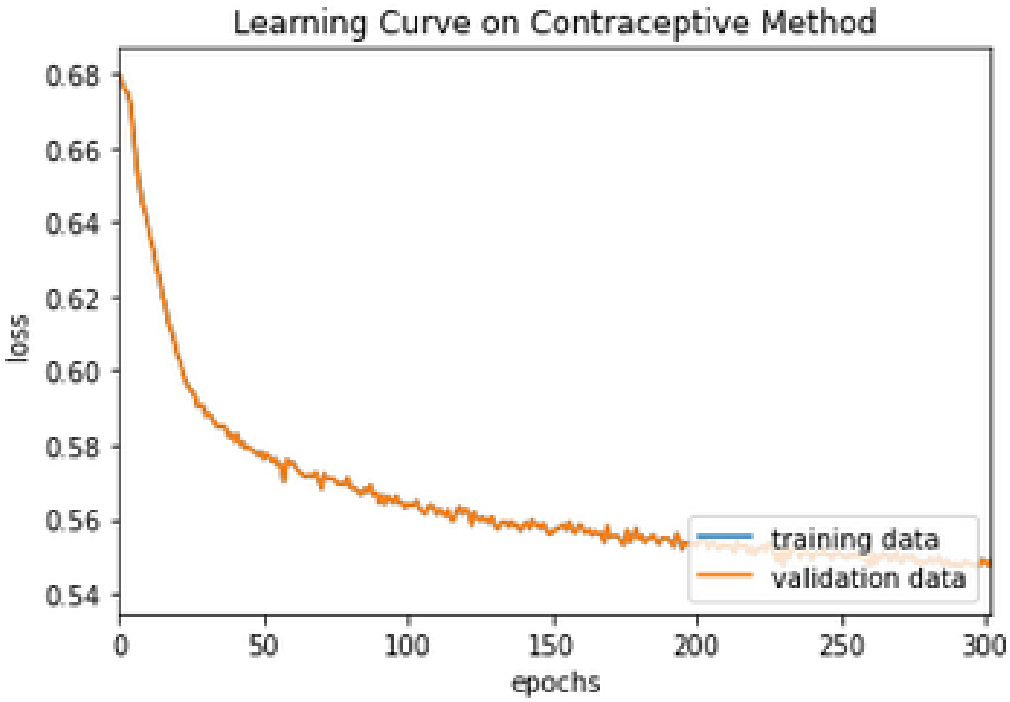

Figure 4: Curve with no over fit

In figure 5 we see that the inputs quickly decrease with no sign of increasing. from this data we can see a significant case of over-fitting.

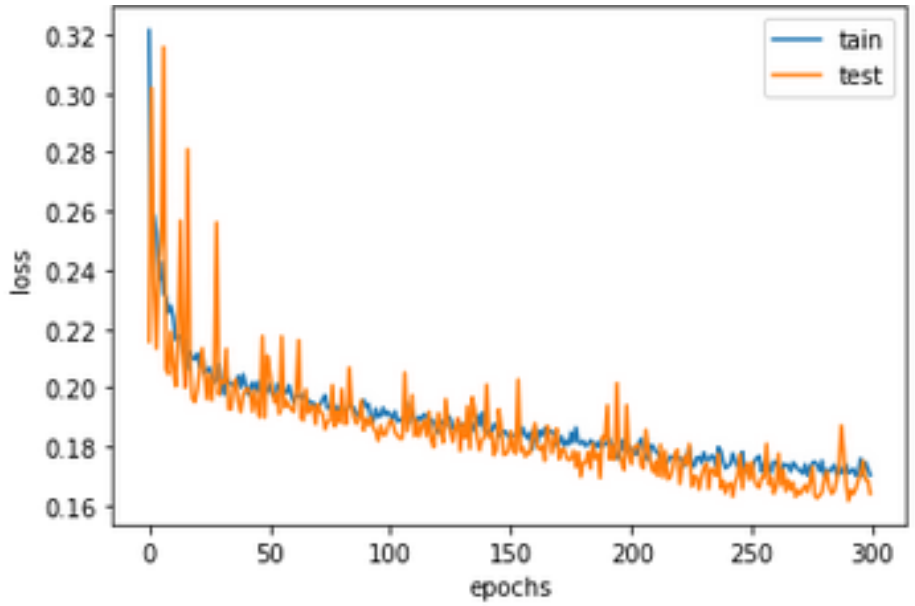

Figure 5: Curve with over fit

Different Neural Networks

Testing out different values on this data base has given me the prediction that most women do not use any kind of contraceptive and that's why the family number is large (number of children) due to many life factors that makes them decide on that. However the majority use no contraceptive.

As this is a Multi-Class Classification instead of the use of stigmoid we used soft max which is a method that squeeze's the data set between 0-1. Our loss is a categorical cross entropy where each example belongs to a single class.

Investigating onto the data, we see in the figures below how each one is when for example Figure 4 is a single layer neuron our activation was a stigmoid and changed our accuracy percentage to around 76% whereas, our multi layer neuron we ran it under the soft max activation and gave us results of 90%. Which indicates that our data set best fits with the multi layered neural network.

```
16/16 [==============================] - 0s 1ms/step - loss: -132.3545 - accuracy: 0.0930
Epoch 97/100
16/16 [==============================] - 0s 1ms/step - loss: -133.4888 - accuracy: 0.0930
Epoch 98/100
16/16 [==============================] - 0s 2ms/step - loss: -134.6134 - accuracy: 0.0930
Epoch 99/100
16/16 [==============================] - 0s 2ms/step - loss: -135.7440 - accuracy: 0.0930
Epoch 100/100
16/16 [==============================] - 0s 2ms/step - loss: -136.8716 - accuracy: 0.0930
Time elapsed: 3.374675989151001s
Accuracy on test set-0.07610993657505286
```

Figure 6: Single Layer Neuron

```
Epoch 97/100
16/16 [==============================] - 0s 1ms/step - loss: 3.7479e-07 - accuracy: 0.0670
Epoch 98/100
16/16 [==============================] - 0s 1ms/step - loss: 3.7479e-07 - accuracy: 0.0670
Epoch 99/100
16/16 [==============================] - 0s 1ms/step - loss: 3.7479e-07 - accuracy: 0.0670
Epoch 100/100
16/16 [==============================] - 0s 2ms/step - loss: 3.7479e-07 - accuracy: 0.0670
Time elapsed: 3.249220609664917s
Accuracy on test set-0.09090909090909091
```

Figure 7: Multi-Layer Neuron

5 Feature importance & reduction

Feature importance

Feature selection or importance is basically the process of reducing the number of features in a data set without losing important information. In our case, the data set haven't changed we still have 6 important features in this data set. As you can see in Figure 7 below all of the features are present and each bar is at a different rate. The number of children has the highest accuracy, observing my data base i realized that it is heavily dependant on the number of children and that's the fact that women who don't take any contraceptive method have a larger number of children in the household. Data set being dependant on that feature gives us an accuracy of 55 % overall. If we remove number of children our overall percentage would be higher.

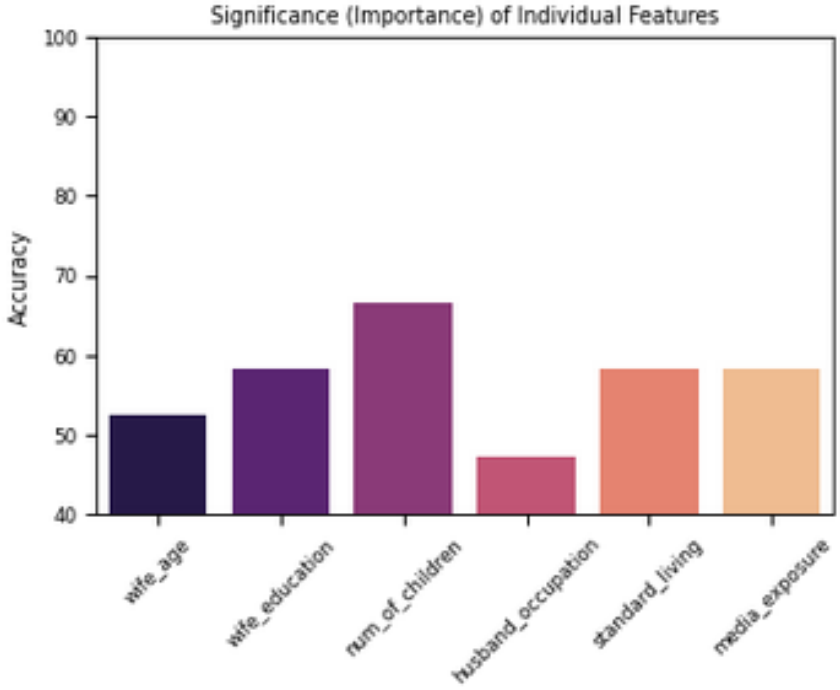

Figure 8: Feature Importance

Feature Extraction

Feature reduction gives us fewer resources to finish computations of data set. Less computation time and less storage capacity gives the time to do more work. As well as it makes data easier to visualize for humans.
After running code removal:
Iteration 0: features
$['wife_age', 'wife_education', 'num_of_children', 'husband_occupation', 'standard_living', 'media_exposure']$
$Iteration 1 : features$
$['wife_age', 'wife_education', 'num_of_children', 'standard_living', 'media_exposure']$
$Iteration 2 : features$
$['wife_education', 'num_of_children', 'standard_living', 'media_exposure']$
$Iteration 3 : features$

$['num_of_children','standard_living','media_exposure']$
$Iteration4 : features$
$['num_of_children','media_exposure']$
$Iteration5 : features$
$['num_of_children']$.

These steps showed the most important feature and pops it out of the rest. Each feature it compared to each other in my case it broke it down to the number of children which is the same at our first prediction to this data set. The iteration organizes my features in the next section.

After Extracting Features

Extracting features derive information from the feature set to build a whole new structure.In figure 8 it shows how each feature is after removal and it changed significantly, the first bar shows the whole average of the features and then it changed the position of each feature due to the results applied and looking closely the feature with the highest accuracy rate after the removal is the wives education. Why is that? women who had a higher education degree were low due to the way of living which makes the majority of the population with lower education degree (high school -max) not educated enough and that contributes on why we have more number of children in the household.

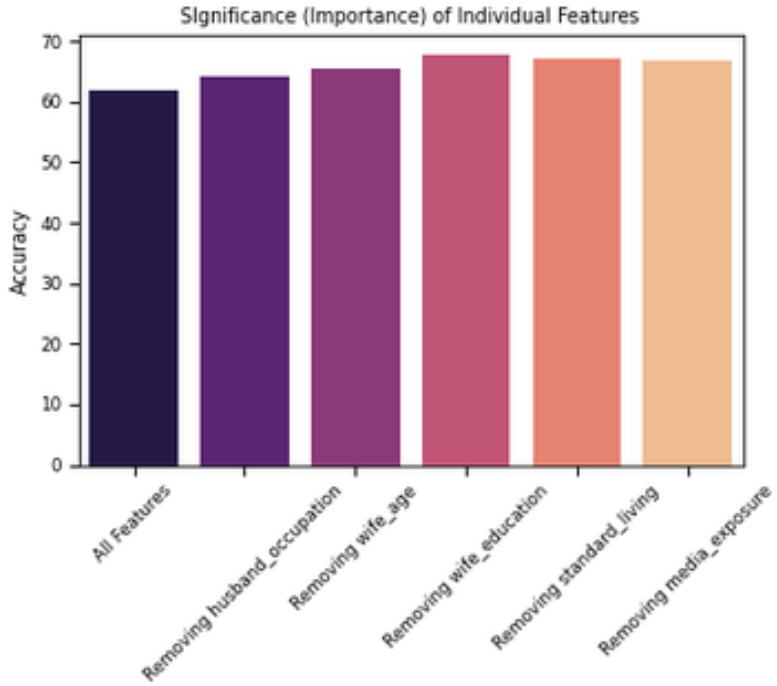

Figure 9: Removal of Features

6 The End

Conclusion

Testing my models has shown me how each one is affected by different factors. Our data improved but not by a whole lot, I noticed that our data is around 53%- 54% and that's because the number of children has a bigger percentage of all our other features. Other features affect on each other, due to number of children being high the number of wives taking contraceptive methods are low.
This report taught me how to deal with a specific data set and work on it under different circumstances such as over-fitting, learning curve, early stopping etc..