# MM 811 Assignment 3

## Description of the Data Set

This data set is imported from the the UCI Machine Learning Repository, which contains descriptions (22 attributes) of 23 species of gilled mushrooms in the Agaricus and Lepiota Family, where each mushroom is classified as either edible, or poisonous.
(Total of 8124 instances)

Obtain the data set from the UCI Machine Learning Repository link:
https://archive.ics.uci.edu/ml/datasets/Mushroom (Save the dataset as a .csv)
After obtaining the data set, change all letters to numeric values. (0 to (n-1), where n is the total number description for that attribute). However, the data set with the numeric values is already provided on the github repository.

## Description of Problem

The problem is: with the 22 attributes, determine if the mushroom is poisonous (0) or edible (1).

## Description of Conditioned Inputs

Since there are 22 attributes (as mentioned in the description of the data set), the inputs consists of the second column to the last column of the mushroom data set. (A matrix of 8124x22)
The outputs is the first column of the mushroom data set. (A matrix of 8124x1)

## Interpretation of My Outputs

Based on the results from my outputs, more hidden layers, lead to better results. More neurons in the hidden layers will also lead to better results. However, if the number for the last hidden layer (before the output) is higher or lower than the number of outputs, the results would not be as good.

## Performance of each Architecture

| Architecture # | Accuracy | TP | TN | FP | FN |
|---|---|---|---|---|---|
| 1 | 0.9557 | 415 | 362 | 22 | 14 |
| 2 | 0.9926 | 424 | 383 | 1 | 5 |
| 3 | 1.0000 | 429 | 384 | 0 | 0 |

## Possible Ways of Improving the Result

As I have found, that one of the architectures have given me really good results (100%), if there are many hidden layers with the first few hidden layers having more neurons, and the number for the last few hidden layers not being less than the number of outputs, the results should improve.

## Summary

With the 22 attributes provided by the mushroom data set and deep learning, most mushrooms can be classified correctly as either poisonous or edible. The results are improved as more hidden layers and number of neurons are added, and that the number of neurons for the last few hidden layers is not less than the number of outputs.

## References

**https://github.com/abramhindle/theanets-tutorial**

**https://archive.ics.uci.edu/ml/datasets/Mushroom**