# Homework 2

Your Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Student ID:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**1.** [40] Use Naïve Bayes Classifier to classify the objects

We conducted a survey to collect people’s daily diets and try to build a model to predict whether their diets result in healthy conditions or not. The final results could be Yes, No, Unsure

|  |  |  |  |
| --- | --- | --- | --- |
| **Breakfast** | **Lunch** | **Dinner** | **Healthy?** |
| Ham | Carnivorous | Beef | Y |
| Milk | Carnivorous | Beef | N |
| Bread | Veggie | Pork | U |
| Bread | Veggie | Veggie | Y |
| Ham | Veggie | Veggie | Y |
| Bread | Carnivorous | Beef | N |
| Ham | Veggie | Pork | N |
| Milk | Veggie | Pork | U |
| Milk | Carnivorous | Veggie | U |
| Noddle | Carnivorous | Pork | ? |

1). [15 points] What is laplace smoothing? And why we need it in the Naïve Bayesian classifier?

2). [25 points] Using the Naive Bayesian Classification Hint: you may need to use laplace smoothing (use the formula in our slide) if you do have zero-conditional probabilities. Use the setting in the slide to solve the problems in this case. Note, only apply laplace smoothing to the ones you have zero-conditional probabilities.

**2. (60 points) Python practice for Naïve Bayes classification**

**Use the Loans data, and run 5 Naïve Bayes techniques to find the best parameters and performance**

* Use Loans\_20K.csv data by using 10-fold cross validation
* Use Loans\_200K.csv data by using 75% as training, 25% as testing

Note:

* You need to change different/multiple parameters to find the best NB model.
* You can find data sets from “slide & data” on blackboard system

Submission

* The ipynb and saved html files
* A comparison of different parameters and accuracy values