

Wei Tang
CSE353
Assignment 4
Naïve Bayes Classifier

1. Explain how you handled continuous-valued attributes and missing attributes values in the data set

For Continuous attributes, I calculate the mean and standard deviation of the attributes over training data and store it in to my probability table

When calculate the posterior belief for $p(x|C)$ I would calculate the normal distribution using the value of testing data attribute and the mean and standard deviation in my training data.

Also, To avoid attribute for having a 0 frequency, I initialize the count as 1 instead of 0

For missing attribute I simply just delete it from the dataset

2. Evaluation:

I slice the data in to 5 sets and each time I feed 1 set as testing and other 4 as training data.

My program would print the probability table to terminal and the accuracy per fold and the average accuracy

Next page is the screenshot of my terminal output

```
Accuracy of fold number: 1    82.8282828283 %
Accuracy of fold number: 2    82.793451759 %
Accuracy of fold number: 3    82.8166724718 %
Accuracy of fold number: 4    82.5496342738 %
Accuracy of fold number: 5    82.1781028678 %
Average accuracr: 82.6332288401 %
```

```
Naive -- -bash -- 136x33
get_info
24705882352941176, 0.0020422318125770654), 'Mexico': (0.004352941176470588, 0.023620530209617754), 'El-Salvador': (0.000588235294117647, 0.004200061652281134), 'India': (0.004823529411764706, 0.0018881011097410605), 'Puerto-Rico': (0.0015294117647058824, 0.004970715166461159), 'China': (0.0038823529411764705, 0.0025046239210850803), 'Yugoslavia': (0.000588235294117647, 0.0003467940813810111), 'United-States': (0.9318823529411765, 0.9092555487053021), 'Trinidad&Tobago': (0.00011764705882352942, 0.0009633168927250308), 'Greece': (0.0029411764705882353, 0.0012715782983970408), 'Japan': (0.003411764705882353, 0.0025046239210850803), 'Taiwan': (0.0024705882352941176, 0.0008091861898890259), 'Nicaragua': (0.00011764705882352942, 0.0014257090012330456)}, 'education-num': {'Yes': (11.566117647058823, 2.3393642803520045), 'No': (9.630856966707768, 2.4248372005468277)}, 'capital-loss': {'Yes': (194.11011764705881, 593.6735971172061), 'No': (58.0963316892725, 324.2978241838662)}, 'education': {'Masters': (0.122, 0.03641337854500617), 'Prof-school': (0.04858823529411765, 0.005278976572133169), '12th': (0.0038823529411764705, 0.016684648581997533), 'Assoc-voc': (0.04435294117647059, 0.04165382244143033), '1st-4th': (0.000588235294117647, 0.006203760789149198), 'Assoc-acdm': (0.03964705882352941, 0.03410141800246609), 'HS-grad': (0.21470588235294116, 0.374884401972873), 'Bachelors': (0.29094117647058826, 0.12580918618988904), '9th': (0.0038823529411764705, 0.01807182490752158), '5th-6th': (0.0015294117647058824, 0.01529747225647349), 'Some-college': (0.17988235294117647, 0.22275739827373614), '11th': (0.00811764705882353, 0.04535295930949445), '10th': (0.006705882352941176, 0.031327065351418), 'Doctorate': (0.030705882352941177, 0.004508323057953144), 'Preschool': (0.000588235294117647, 0.0020422318125770654), '7th-8th': (0.005764705882352941, 0.020229654747225646)}, 'fnlwgt': {'Yes': (191597.00705882354, 101077.38392006017), 'No': (191480.93048705303, 110336.21497093348)}, 'marital-status': {'Separated': (0.007176470588235294, 0.03780055487053021), 'Widowed': (0.009058823529411765, 0.033947287299630084), 'Divorced': (0.057529411764705884, 0.1621840320591862), 'Married-spouse-absent': (0.007176470588235294, 0.015143341553637484), 'Never-married': (0.06364705882352942, 0.4083307644882861), 'Married-AF-spouse': (0.0015294117647058824, 0.0006550554870530209), 'Married-civ-spouse': (0.8547058823529412, 0.34220869297163997)}, 'occupation': {'Farming-fishing': (0.013764705882352941, 0.038262946979038225), 'Armed-Forces': (0.000588235294117647, 3.8532675709001234e-05), 'Craft-repair': (0.11164705882352942, 0.14122225647348952), 'Protective-serv': (0.025529411764705884, 0.021154438964241677), 'Transport-moving': (0.04670588235294117, 0.05475493218249075), 'Prof-specialty': (0.2368235294117647, 0.09282521578298397), 'Sales': (0.1356470588235294, 0.11517416769420469), 'Exec-managerial': (0.24576470588235294, 0.09205456226880394), 'Handlers-cleaners': (0.012823529411764706, 0.06076602959309495), 'Adm-clerical': (0.07494117647058823, 0.13505702836004932), 'Other-service': (0.018470588235294117, 0.13366985203452528), 'Tech-support': (0.04341176470588235, 0.03410141800246609), 'Priv-house-serv': (0.000588235294117647, 0.005895499383477189), 'Machine-op-inspct': (0.03494117647058823, 0.07556257706535142))}

Accuracy of fold number: 1 82.0282828283 %
Accuracy of fold number: 2 82.793451759 %
Accuracy of fold number: 3 82.8166724718 %
Accuracy of fold number: 4 82.5496342738 %
Accuracy of fold number: 5 82.1781028678 %
Average accuracy: 82.6332288401 %
Weis-MacBook-Pro:Naive weitangs
```