

Bayesian Cancer Decider

In this worksheet, you will create a Bayesian decider for cancer data. In the end, you will produce a classifier that is able to guess if a particular sample is malignant AND an estimate for how accurate your classifier is.

Input: Cancer Data

The input to your program will be the file named `cancer.csv` in Schoology. The file contains labeled data with three columns of measurement data about a potentially cancerous sample (smoothness, area, radius) and with the final column containing the label (benign or malignant). In total there are 499 data points.

Output: A “Trained” Classifier and an Accuracy Estimate

Your program will produce two things. First, it will create a classifier that is able to take the three measurements present in the data of a potentially cancerous sample and produce a guess of whether the sample is benign or malignant. Then, once the classifier has been “trained”, you will use a test set to determine how accurate the classifier is.

To create the classifier, you will essentially “filter” for all elements of the data set that match a given sample. You will then determine the probability that the sample is benign or malignant. You will use the first 400 rows of the data set to train the classifier.

After you have created the classifier, you will use the remaining 99 rows to “test” how accurate your classifier is. For each element of the test set, you will compare the classifier answer with the label. You will then calculate the percentage that your classifier got correct.

Final Results:

Assuming you did everything correctly, you should get an accuracy for your classifier of around 96%.