DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Homework Assignment No. 07:

HW No. 07: Information Theory and Statistical Significance

submitted to

Professor Joseph Picone
ECE 8527: Introduction to Pattern Recognition and Machine Learning
Temple University
College of Engineering
1947 North 12th Street
Philadelphia, Pennsylvania 19122

April 18, 2024

prepared by:

Leo Berman
Email: leo.berman@temple.edu

A. TABLE FOR PROBLEM 1

	Train	Dev	Uniform
$\mathrm{H}(x_1)$	06.75	06.75	07.00
$\mathrm{H}(x_2)$	06.79	06.81	07.00
$\mathrm{H}(x_1, x_2)$	13.13	12.60	13.99
$\mathrm{H}(x_1\|, x_2)$	06.35	05.79	06.99
$\mathrm{H}(x_2\|, x_1)$	06.38	05.85	06.99
$\mathrm{I}(x_1, X_2)$	00.41	00.97	00.01

Figure 1: Entropy and Mutual Information

B. TABLE FOR PROBLEM 2

Confidence [%]	Number of Samples	Minimum Error Decrease [%]
80	100	04.54
	500	02.09
	2000	01.06
	5000	01.00
	10,000	01.00
85	100	05.53
	500	02.56
	2000	01.30
	5000	01.00
	10,000	01.00
90	100	06.73
	500	03.15
	2000	01.60
	5000	01.02
	10,000	01.00
95	100	08.43
	500	04.00
	2000	02.04
	5000	01.30
	10,000	01.00

Figure 2: Minimum error required for confidence and numbers of samples

C. SUMMARY

When comparing the train and development set, we can see that the entropies of the individual sets are all the same. What this tells us is that the entropy of the data was slightly lower for the marginals of train and dev while uniform comes out to the expected value of 7. This means that the datasets have a bias towards a few numbers which could potentially make the distribution easier to predict. When we look at their joint entropy we can see that Train and Dev once again have a lower entropy when compared to the Uniform, and this essentially means that the dataset is biased towards certain data points and since the entropy of the Dev set is less than that of the Train set, it is likely that the two sets are correlated. Since uniform is essentially the sum of the two variables before it, it shows that the two variables are likely independent. For the entropy of the variables $H(x_1|Hx_2)$, we can see that the Train and Dev sets once again have a lower entropy than the Uniform set which means that in Train and Dev, the two variables are dependent on each other more than in the uniform set which makes sense as in uniform generation, the two variables are independent. Our observations are reflected in the mutual information as we can see that the two vectors are the least informant on each other in the Uniform set, and most informant on each other in the dev set.

As we can see in Figure 2, the shape of the data is representative to what we expect. As the confidence increases, we either need more samples or a more significant error decrease to prove statistical significance. More specifically, no matter the confidence %, when the number of samples reaches 10,000 datapoints, every system shows statistical significance at a 1 percent error decrease and this could theoretically be shown to be an even lower number if the proper experimentation was done.