

Multimodal sensing

Homework 1

Student: Adil Hydari, adil.hydari@rutgers.edu

Professor: Jorge Ortiz

Problem 1: Mutual Information in Multimodal Systems

- (a) Use the definition of the mutual information equation:

$$I(X; Y) = H(X) - H(X | Y)$$

to explain how mutual information quantifies the reduction in uncertainty about Y when you know X_1 and X_2 .

- (i) Given that X_1 and X_2 are conditionally independent on the target variable Y , the joint mutual information function

$$I(X_1, X_2, Y)$$

is the sum of the individual mutual information terms $I(X_1; Y)$ and $I(X_2; Y)$. This can be derived from first the chain rule for conditional entropy:

$$H(X_1, X_2 | Y) = H(X_1 | Y) + H(X_2 | X_1, Y)$$

However, based on the fact that X_1 and X_2 are independent variables we can say that:

$$H(X_1, X_2 | Y) = H(X_1 | Y) + H(X_2 | Y)$$

Which is to say that the total remaining entropy with knowledge of X_1 and X_2 with respect to Y is the sum of the remaining entropy of X_1 and X_2 , respectively. Finally, we can assert that the resulting mutual information function $I(X_1, X_2; Y)$ when using the chain rule and the conditional independence of X_1 and X_2 the resulting Mutual Information function is:

$$I(X_1, X_2; Y) = I(X_1; Y) + I(X_2; Y)$$

The resulting Mutual Information function shows that the information provided by both modalities X_1 (sensor data) and X_2 (audio data) about the target variable Y (outcome) can be understood as the sum of the information each modality provides individually about Y . In total, both X_1 and X_2 independently contribute to the reduction in entropy which is represented by the function $I(X_1, X_2; Y)$

Problem 2: Mutual Information in Multimodal Systems

- (a) Provide a real-world example in an IoT or smart home environment; for instance, consider the integration of temperature sensor data X_1 and audio data X_2 for monitoring a home's security. Discuss the importance and role of understanding mutual information in such systems, highlighting how it helps measure the contribution of each data stream to the prediction of Y .
- (i) In the case of a smart home security system that integrates streams of audio and temperature data to monitor and predict security events, such as fires. In this case we can have 2 primary ways to detect and predict this fire data:
- X_1 : Temperature sensor data.
 - X_2 : Audio data.

With the eventual goal being to predict the **Fire Event** Y .

- In the case of X_1 , the sensor will be able to detect abnormal temperature spikes which may indicate the generalization error is derived from the Rademacher complexity presence of a fire.
- In the case of X_2 , the sensor will be able to detect glass shattering or the sound of smoke alarms from the presence of a fire.
- Broadly, for Y , we can say that the state of Y falls into 2 categories:
 $Y = 1$ Presence of fire
 $Y = 0$ Normal state

The Mutual information function:

$$I(X_1, X_2; Y) = I(X_1; Y) + I(X_2; Y)$$

tells us that the total reduction in entropy for Y is the sum of the reductions from both the temperature and audio data. Which means that both the audio and temperature sensor data independently contribute to the detection of a fire.

The understanding of Mutual information in this case has a few benefits:

- Prioritizing data streams: By calculating $I(X_1, Y)$ and $I(X_2, Y)$, the engineer can determine which data stream contributes more accurately to predicting security events. For example, if temperature data X_2 provides more information about fires than audio data X_1 , the engineer can prioritize the deployment of accurate temperature sensors over expensive microphones.
- Optimizing system performance: In the case of a Multimodal system that has redundant or overlapping information streams; meaning that the Mutual information from sensor X_y provides redundant data to that of sensor X_z . With the understanding of Mutual information, the engineer can optimize or reduce the amount of sensors needed for a certain task, this can both improve system performance, as well as reducing cost.

Problem 3: Generalization Bound in Multimodal Learning

(a) Explain how the generalization error bound is derived from the Rademacher complexity, which measures the richness of a hypothesis class in terms of its ability to fit random noise. How does increasing the number of modalities X_1, X_2, \dots, X_n increase the sum of the complexities $R(H_1) + R(H_2) + \dots + R(H_n)$, thereby increasing the generalization error? Why must the number of training samples m also increase to control this generalization error?

- (i) The Rademacher complexity $R(H)$ of a hypothesis class H is defined as the expectation of how well the functions in H can fit random noise. Higher complexity implies a wider breadth of hypothesis classes that can potentially overfit the training data. The $1/m$ indicates that increasing the number of training samples reduces the generalization error, in general this can be seen as an increase in training data leading to a better fitting hypothesis class. Finally,

$$\frac{\log(1/\delta)}{m}$$

term accounts for the probability (δ) that the bound holds. A smaller δ (higher confidence) results in a larger bound due to the $\log()$.

- (ii) Each modality X_i comes with its own feature matrix and thus an associated hypothesis class. Further, $R(H_i)$ quantifies how well H_i can fit random labels; In a multimodal scenario, when we add these hypothesis classes together in the form of a Rademacher complexity this increases the risk of overfitting unless the amount of training data is increased. Mathematically:

$$R(H) \leq \sum_{i=1}^n R(H_i)$$

as n increases for the error bound equation, unless the denominator m increases as well, the error bound will increase due to the direct proportionality of m and the summation of the $R(H)$ function. When considering the increase in modalities, we must first go back to the fact that each multimodal data stream has its own hypothesis class and thus its own Rademacher complexity associated with it. Considering what I just laid out, we can draw the conclusion that as we add modalities, we increase the amount of training data to maintain an acceptable error bound. On the other hand, if we add too many modalities, this can create an endless feedback loop in which we constantly end up having to find new data for each different modality to correctly fit the data. Even if we manage to find clean data to correctly fit our data, we still have the issue of training time, when we add more training data m , we end up having to spend more time training in order to correctly fit our model on m .

Problem 4: Synthetic Time Series Data Generation and Mutual Information Calculation

- a Discuss how the temporal dependencies affect the mutual information, particularly in a multimodal context where time series data from different sensors or sources need to be synchronized. Analyze how the strength of the temporal correlation between the variables impacts the mutual information and explain why understanding this relationship is critical in designing multimodal systems.
- (i) When variables have temporal dependencies, the mutual information between them can increase because the variables share more predictable patterns over time. In multimodal systems, aligning the time series data from different modalities makes sure that these temporal dependencies are correctly captured. When A and B have high correlation, knowing the value of one variable reduces the uncertainty of the other, which in turn increases mutual information; this is also applicable vice-versa. Particularly in multimodal systems, data from different sensors must be synchronized in time in order to observe temporal dependencies. Misalignment in the time series can obscure correlation between the variables, thus reducing mutual information. When we consider a smart-home scenario: proper synchronization increases the shared mutual information between the temperature (following the fire detection example) data and the video data. This can be attributed to the accuracy of simultaneous events on sensors, for example if we consider a fire event, a small fire seen underneath a door and a rise in temperature across the whole house happening simultaneously would confidentially trigger a fire event in the multimodal model. Further, if we have properly synchronized and temporally dependent variables, we can avoid incorrect associations between unrelated events.