LECTURE 1

Introduction

2 Handouts

Lecture outline

- Goals and mechanics of the class
- notation
- entropy: definitions and properties
- mutual information: definitions and properties

Reading: Ch. 1, Scts. 2.1-2.5.

Goals

Our goals in this class are to establish an understanding of the intrinsic properties of transmission of information and the relation between coding and the fundamental limits of information transmission in the context of communications

Our class is not a comprehensive introduction to the field of information theory and will not touch in a significant manner on such important topics as data compression and complexity, which belong in a source-coding class

Notation

- random variable (r.v.) : X
- sample value of a random variable : x
- set of possible sample values x of the r.v. X : $\mathcal{X}$
- Probability mass function (PMF) of a discrete r.v. X : $P_X(x)$
- Probability density function (pdf) of a continuous r.v. : $p_X(x)$

Entropy

- Entropy is a measure of the average uncertainty associated with a random variable
- The entropy of a discrete r.v. X is $H(X) = -\sum_{x \in \mathcal{X}} P_X(x) log_2 (P_X(x))$
- entropy is always non-negative
- Joint entropy: the entropy of two discrete r.v.s X, Y with joint PMF $P_{X,Y}(x, y)$ is:

$$H(X,Y) = -\sum_{x \in \mathcal{X}, y \in \mathcal{Y}} P_{X,Y}(x,y) log_2 \left(P_{X,Y}(x,y) \right)$$

- Conditional entropy: expected value of entropies calculated according to conditional distributions $H(Y|X) = E_Z[H(Y|X = Z)]$ for r.v. Z independent of X and identically distributed with X. Intuitively, this is the average of the entropy of Y given X over all possible values of X.

Conditional entropy: chain rule

$$
\begin{aligned}
& H(Y|X) = E_Z[H(Y|X=Z)] \\
= & -\sum_{x\in\mathcal{X}} P_X(x) \sum_{y\in\mathcal{Y}} P_{Y|X}(y|x) log_2[P_{Y|X}(y|x)] \\
= & -\sum_{x\in\mathcal{X},y\in\mathcal{Y}} P_{X,Y}(x,y) \log_2[P_{Y|X}(y|x)]
\end{aligned}
$$

Compare with joint entropy:

$$
\begin{aligned}
& H(X,Y) \\
= & -\sum_{x\in\mathcal{X},y\in\mathcal{Y}} P_{X,Y}(x,y) \log_2[P_{X,Y}(x,y)] \\
= & -\sum_{x\in\mathcal{X},y\in\mathcal{Y}} P_{X,Y}(x,y) \log_2[P_{Y|X}(y|x)P_X(x)] \\
= & -\sum_{x\in\mathcal{X},y\in\mathcal{Y}} P_{X,Y}(x,y) \log_2[P_{Y|X}(y|x)] \\
& -\sum_{x\in\mathcal{X},y\in\mathcal{Y}} P_{X,Y}(x,y) \log_2[P_X(x)] \\
= & H(Y|X) + H(X)
\end{aligned}
$$

This is the **Chain Rule** for entropy:

$H(X_1,\ldots,X_n) = \sum_{i=1}^{n} H(X_i|X_1 \ldots X_{i-1})$. Question: $H(Y|X) = H(X|Y)$?

Relative entropy

Relative entropy is a measure of the distance between two distributions, also known as the Kullback Leibler distance between PMFs $P_X(x)$ and $P_Y(y)$.

Definition:

$$D(P_X||P_Y) = \sum_{x \in \mathcal{X}} P_X(x) \log\left(\frac{P_X(x)}{P_Y(x)}\right)$$

in effect we are considering the log to be a r.v. of which we take the mean (note that we assume $0 \log(\frac{0}{p}) = 0$ and $p \log(\frac{p}{0}) = \infty$

Mutual information

Mutual Information: let X, Y be r.v.s with joint PMF $P_{X,Y}(x, y)$ and marginal PMFs $P_X(x)$ and $P_Y(y)$

Definition:

$$
\begin{aligned}
& I(X;Y) \\
= & \sum_{x\in\mathcal{X},y\in\mathcal{Y}} P_{X,Y}(x,y)\log\left(\frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)}\right) \\
= & D\left(P_{X,Y}(x,y)||P_X(x)P_Y(y)\right)
\end{aligned}
$$

intuitively: measure of how dependent the r.v.s are

Useful expression for mutual information:

$$
\begin{aligned}
& I(X;Y) = H(X) + H(Y) - H(X,Y) \\
= & H(Y) - H(Y|X) \\
= & H(X) - H(X|Y) \\
= & I(Y;X)
\end{aligned}
$$

Question: what is $I(X;X)$?

Mutual information chain rule

Conditional mutual information: $I(X;Y|Z) = H(X|Z) - H(X|Y,Z)$

$$
\begin{aligned}
& I(X_1, \ldots, X_n; Y) \\
= \;& H(X_1, \ldots, X_n) - H(X_1, \ldots, X_n|Y) \\
= \;& H(X_1, \ldots, X_n) - H(X_1, \ldots, X_n|Y) \\
= \;& \sum_{i=1}^{n} H(X_i|X_1 \ldots X_{i-1}) \\
& - \sum_{i=1}^{n} H(X_i|X_1 \ldots X_{i-1}, Y) \\
= \;& \sum_{i=1}^{n} I(X_i; Y|X_1 \ldots X_{i-1})
\end{aligned}
$$

Look at 3 r.v.s: $I(X_1, X_2; Y) = I(X_1; Y) + I(X_2; Y|X_1)$ where $I(X_2; Y|X_1)$ is the extra information about Y given by X_2, but not given by X_1