Applied Information Theory

Final Review Sheet

1 Overview

For the final you should cover all of the following:

- The items on the midterm review.
- The items on this review.
- All of the reading assignments from Cover and Thomas that are posted on the course web page (in the rightmost column of the table).
- All of the problem sets, especially number 5, for which solutions were handed out in class on the last day of class.

2 Notation

On this review sheet and the exam, like on the last review sheet and in the assignments, I will use the following notation, unless specified otherwise:

- Random variables in capitals: X, Y, Z, X_1, X_2.
- Sets representing outcomes of a random variable in calligraphic upper case: $\mathcal{X} = \{1, 2, 3, 4, 5, 6\}$.
- Sample value of a random variable in lower case x, y.
- Multiple samples of a random variable: $x_1, x_2, ..., x_n$.
- Generic discrete probability distribution $p(x)$.
- Generic continuous probability density $f(x)$.
- Cumulative distribution function of a density $f(x)$: $F(x)$.
- Discrete entropy: $H(X)$ or $H(p(x))$, or for a Bernoulli random variable $H(p)$.
- Differential entropy: $h(X)$ or $h(f(x))$.
- KL-divergence between probability distributions: $D(p(x)||q(x))$ or $D(p||q)$.
- Mutual information between random variables X and Y: $I(X;Y)$.
- Mutual information between random variable X and the joint distribution of Y and Z: $I(X;Y,Z)$.

3 Specific material

- Typicality. Know the exact mathematical definition. Know the approximate size of the typical set, and the asymptotic probability of a sample being in the typical set.
- Joint typicality. Know the exact mathematical definition of joint typicality. There are three parts of joint typicality: typicality of X, typicality of Y, and typicality of the pair (X,Y). Make sure you understand how a set of pairs could be typical with respect to the joint distribution $p(x, y)$ even when the set of x's is not typical, and vice versa.
- What does the AEP state? That the negative of the average log probability of a sample converges to the entropy (which is the expectation of the negative of the average log probability) as n goes to infinity.
- Kraft Inequality. Know the Kraft inequality.
- Know the meaning of the following types of source codes: uniquely decodable, prefix codes, instantaneous codes (note, the last two are two names for the same thing).
- Using the "wrong code". If you build a source coding scheme using a distribution $q(x)$ when the data was distributed as $p(x)$, it will take you an additional number of bits to encode each symbol that is equal to $D(p||q)$.
- Shannon coding. Shannon coding is the type of coding described in class (and in the book) in which you have one mechanism for coding typical sequences, and another method for coding non-typical sequences. You should be able to argue that for long sequences, this coding scheme will allow you to code at rates approaching the entropy of the source.
- Formal specification of a channel. A channel is defined by an input alphabet $\mathcal{X}$, a conditional distribution of outputs given inputs ($p(y|x)$) and an output alphabet $\mathcal{Y}$.
- Definition of channel capacity. The definition I want you to know is

$$\max_{p(x)} I(X;Y).$$

- Examples of channels: noiseless, noisy typewriter, BSC, BEC. These are all in the book. Know what they are and how they behave. That is, how do they affect codewords that are sent over them?
- Rate of an (M, n) code. Know the formula, i.e.

$$R = \frac{\log_2 M}{n}.$$

- Huffman codes. Know what they are, and how to build one.
- Basic outline of the forward part of the channel coding theorem. You should know
  - To prove it, you compute the average error over both codewords and codebooks.
  - The codebooks are built by randomly sampling from the distribution $p(x)$.
  - Because the codewords are all built "symmetrically", one can compute the average error by focusing on a particular codeword, in particular, the first codeword, and averaging over all codebooks.
  - In jointly typical decoding, which is used in the theorem, an error occurs when either, a) the output is not jointly typical with the sent codeword, or b) the output *is* jointly typical with something other than the sent codeword.

  - The jointly typical decoding error is small on average, so there must be at least one codebook for which the error is less than or equal to the average.
  - The average error over codewords in the optimal codebook is small, so the best half of the codewords must have error that is small.
  - Thus, there exists a codebook and a set of codewords that all have small error.

- Be able to derive the capacity for a binary symmetric channel.
- Know the difference between "detecting an error" and "correcting an error" in channel coding.
- Know how to build a Hamming code, how to use it to send codewords, and how to do error corrections as done in Assignment 6.