Telecom Paris ACCQ204, Coding Theory

Assignment 1 - Solutions

Exercise 1 (RAID, distributed storage)**.** Redundant Arrays of Independent Disks consist of a set of disks such that any subset of s disks can be disabled and the others are still able to reconstruct any requested file (the system can tell which disks are disabled).The rate of a RAID system corresponds to the rate at which data is stored.

1. Design a RAID system for 7 disks and $s = 2$. To do this you may want to consider the $(7, 4)$ Hamming code.
2. What happens if we use this code and try correct 3 erasures?

Solution.

1. Encode each successive 4 bits of data into the corresponding seven bit codeword of a $(7, 4)$ Hamming code, and write each bit of the codeword on a different disk. If s disks are disabled, this means that we get to observe codewords with erasures at some specific s positions. To reconstruct the original codeword x from a corrupted (i.e., x with 2 erased positions) vector y, observe that because the minimum distance is 3, all except codeword x will differ from y in at least one of the non-erased positions. Therefore only x will be consistent with y, and erasure decoding will be error-free.
2. Things might go wrong. Consider two codewords whose distance is 3 and suppose one of them is stored. If the 3 erased positions correspond to the positions where the 2 codewords differ then it won't be possible to (fully) recover the original data.

□

Exercise 2. Let C be a code with minimum distance d. Prove that C can correct any pattern of e_1 errors and e_2 erasures provided that $2e_1 + e_2 + 1 \leq d$. (Hint: given an erasure pattern, consider the code obtained by the deleting the erasure positions.)

Solution. Consider a pattern of $e_2 \leq d - 1$ erasures, and the code obtained by deleting the erasure positions of the code. The resulting code C' has a minimum distance at least $d - e_2$ and thus can be corrected as long as $2e_1 \leq (d - e_2) - 1$. Once C' has been error corrected, C can be erasure corrected since $e_2 \leq d - 1$ (see Ex.1).

□

Exercise 3 (Best decoder)**.** Consider a set of $\mathcal{M}$ messages. A random message M is chosen with probability $P(M = m) = p_m$ (hence $\sum_m p_m = 1$), encoded, and sent across a channel. Upon observing the channel output y, the receiver declares one of the messages by means of a decoder which maps each channel output to one of the messages. Let D^* be the Maximum A Posteriori (MAP) decoding rule, i.e.

$$D^*(y) = \arg\max_m P(m|y).$$

1. Show that among all decoding functions, D^* minimizes the error probability given any channel output.
2. Deduce that D^* minimizes the average error probability among all decoding function.

Solution. 1. If a channel output y is decoded into message m, $P(error|y) = 1 - P(m|y)$. Therefore D^* minimizes the error probability conditioned on y.

2. Since $P(error) = \mathbb{E}(P(error|Y))$, D^* also minimizes the average error probability.

□

Exercise 4 (MAP decoder)**.** Consider communication over a binary symmetric channel with crossover probability p. There are two possible equally likely messages that are encoded over three bits: 000 and 111. What is the error probability of the MAP decoder?

Solution.

$$P(error) = 3p^2(1-p) + p^3 \tag{1}$$

□

Exercise 5 ($A(n, d, w), A(n, d)$)**.** For any integers n, d, w with $d \leq 2w \leq n$, let $A(n, d, w)$ be the largest possible size of a set of binary vectors of length n and weight w whose minimum distance is at least d, and let $A(n, d)$ be the largest possible size of a set of length n binary vectors whose minimum distance is at least d. Prove that

$$A(n, d) \leq \sum_{w=0}^{n} A(n, d, w)$$

Solution. Consider a code which achieves $A(n, d)$. This code is the disjoint union of classes of codewords of different weight w. Since each class has a minimum distance at least equal to d it has at most $A(n, d, w)$ elements. □