Proof of

The Noisy-Channel Coding Theorem
Information theory, pattern recognition, and neural networks

- 1. Noisy-channel coding
- 2. Source coding (Data compression)
  - 2.1 Information content, entropy
  - 2.2 Typicality and the source coding theorem
  - 2.3 Symbol codes
  - 2.4 Symbol codes and Arithmetic coding
- 3. Noisy-channel coding
  - 3.1 Inference and Information measures for noisy channels
  - 3.2 Capacity of a noisy channel
  - 3.3 The noisy-channel coding theorem
Definition of Capacity

The Capacity of a channel is the maximum, over all input distributions $P(x)$, of the mutual information:

$$C \equiv \max_{P_X} I(X; Y)$$

The distribution $P_X^*$ that achieves the maximum is called the optimal input distribution.

Shannon's noisy channel coding theorem:

Reliable (virtually error-free) communication is possible at rates up to $C$.
Noisy typewriter

\[ P(y = F \mid x = G) = \frac{1}{3}; \]
\[ P(y = G \mid x = G) = \frac{1}{3}; \]
\[ P(y = H \mid x = G) = \frac{1}{3}; \]
\[ \vdots \]

- What is the optimal input distribution?
- What is the capacity?
Capacity = \max P_x \ I(X; Y)

= \max P_x \ H(Y) - H(Y|X)

= \log 27 - \log 3

= \log 9 \text{ bits}
Non-confusabe subset of inputs

A
B
C
D
E
F
G
H
I
...  
Y
Z
Z

reliable communication is possible at $C = \log_2 9$ bits
\[ P^*_X = \{ 2, \frac{1}{27}, \frac{1}{27}, \frac{1}{27}, \ldots, \frac{1}{27} \} \]

\[ P^*_X = \{ 3, 0, \frac{1}{9}, 0, 0, \frac{1}{9}, 0, 0, \ldots \} \]
Extended channel

$g_{BSC}$

$N$ uses

$fN$ flips

$\pm \sqrt{N}$
almost certainly

\[ \#\text{non-confusible inputs} \]

\[ \frac{2^{NH(Y)}}{2^{NH(Y|X)}} = 2 \]

\[ NI(X;Y) \]
infusable inputs

\[
\frac{NH(y)}{2NH(y|x)} \leq 2 \leq 2NC \rightarrow NC \text{ bits per } N
\]
almost certainly
# non-observable inputs

\[ NHC(X) \]

\[ NHC(Y) \]

\[ NHC(Z) \]
For any channel:
Reliable (virtually error-free) communication is possible
at rates up to $C$
For the BSC with flip prob \( f \),

(whose capacity is)

\[ C = 1 - H_2(f) \]

for any \( \epsilon > 0 \) \& \( R < C \)

for large enough \( N \)

\[ \exists \text{ a code of length } N \text{ \& rate } R \geq R \]

and a decoder s.t. the probability of block error is \( < \epsilon \)
(7,4) Hamming Code

\[ H = \begin{bmatrix}
1 & 1 & 1 & 0 & 1 & 0 & 0 \\
0 & 1 & 1 & 1 & 0 & 1 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 1
\end{bmatrix} \]

\[ M = 3 \]

\[ N = 7 \]

Valid transmissions \( t \) satisfy

\[ H t = \begin{bmatrix}
0 \\
0 \\
0
\end{bmatrix} \mod 2 \]
(7, 4)

Syndrome

deduce
most probable explanation
(7,4) Hamming Code

\[
H = \begin{bmatrix}
1 & 1 & 1 & 0 & 1 & 0 & 0 \\
0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 \\
1 & 0 & 1 & 1 & 1 & 0 & 0 & 1
\end{bmatrix}
\]

\[M = 3\]

\[N = 7\]

Valid transmissions \(t\) satisfy

\[
Ht = \begin{bmatrix}
0 \\
0 \\
0
\end{bmatrix} \mod 2\]
Parity check matrix

\[
H = \begin{bmatrix}
1 & 1 & 1 & 0 & 1 & 0 & 0 \\
0 & 1 & 1 & 1 & 1 & 1 & 0 \\
1 & 0 & 1 & 1 & 1 & 0 & 1 \\
\end{bmatrix}
\]

\[
K = 4 \\
N = 7
\]
(7,4) Hamming Code

\[
H = \begin{bmatrix}
1 & 1 & 1 & 0 & 1 & 0 & 0 \\
0 & 1 & 1 & 1 & 0 & 1 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 1
\end{bmatrix}
\]

\[M = 3\]

\[N = 7\]

Valid transmissions \( t \) satisfy

\[
H t = \begin{bmatrix}
0 \\
0 \\
0
\end{bmatrix} \mod 2
\]

Received signal \( r = t + n \)
Syndrome \( z = Hr = Hn \).
Syndrome decoder \( z \rightarrow \hat{n} \).
flip bit 2

t = 0 0 0 0 0 0 0 0
r = 0 1 0 0 0 0 0 0
Parity check matrix

\[ H = \begin{bmatrix} 1 & 1 & 1 & 0 \ 0 & 1 & 1 & 1 \ 1 & 0 & 1 & 1 \ 1 & 0 & 0 & 1 \end{bmatrix} \]

\[ K = 4 \]
\[ N = 7 \]

\[ z = Hr = H(t+n) = Hr + Hn = Hn \]

\[ r = t + n \mod 2 \]
flip bit 2

\[
t = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}
\]

\[
z = \begin{pmatrix} 1 \\ 0 \end{pmatrix}
\]
\[ z = H r \]

\[ \hat{n} = H n \]
\[ K = N - M \]

\[ R = \frac{K}{N} \]
How we won the bent coin lottery

Probability of '1' = $f$

To have a 99.99% chance of winning, we bought all the typical tickets
How we won the bent coin lottery

To have a 99.99% chance of winning, we bought all the typical tickets

Number of tickets in 'typical set'

$|T| \approx 2^{NH_2(f)}$
How to prove good codes exist

Constructive proof

Given required $R < C$, and $\epsilon > 0$,

$$
H = \begin{bmatrix}
1 & 1 & 1 & 0 & 0 & 1 & 0 \\
0 & 1 & 1 & 1 & 0 & 0 & 1 \\
1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\end{bmatrix}
$$

Non-constructive proof
If average weight of all babies is $< \varepsilon$, there must be (at least!) one baby with weight $< \varepsilon$. 
The Capacity of a channel

is the maximum, over all input distributions $P(x)$, of the mutual information:

$$C \equiv \max_{P_X} I(X;Y)$$

The distribution $P_X^*$ that achieves the maximum is called the optimal input distribution.

Shannon's noisy channel coding theorem:
Reliable (virtually error-free) communication is possible at rates up to $C$
Proof for the Binary Symmetric Channel

- Linear block code
- Syndrome decoder
- Use **typical** noise vectors
- Compute **average** probability of error
How we won the bent coin lottery

To have a 99.99% chance of winning, we bought all the typical tickets.

Number of tickets in 'typical set'

\[ |T| \approx 2^{NH_2(f)^+} \]

Typical-set Syndrome Decoder

\[ n \]

\[ 001000001000, \ldots \]

\[ 00 \]

\[ Z = H_n c^{10} \]

\[ M \]
\[ r = t + n \]
\[ z = Hr \]
\[ z = Hn \]
\[ \hat{n} \]

\[ H = \]

\[ N \]

\[ K = N - M \] (5\text{th} \text{M} \text{bits})

\[ t, t_2, t_3, \ldots, t_k, t_{k+1}, \ldots, t_{N} \]

\[ Ht = 0 \]
Probability of error of this code w/ $H$ = $P_{\bar{H}} = P_{II}$ + $P_{II}$

$P_{\bar{H}}$ is that $\bar{H}$ is not in the bag, independently of $H$.

$\bar{H}$ is in the bag, but other $\bar{H}$ are also in the bag w/ $H \bar{H} = Z$

$H(n-\bar{n}) = 0$
\[ T = 2^{NH_2(f)^+} \rightarrow 0 \text{ by picking } N \text{ large} \]

\[ \text{Prob that } N \text{ is not in the bag} \]

\[ \text{indep of } H \]
\[ P_\Pi (H \equiv) = \sum_{n \in \text{bag}} P(n) 1(\exists \tilde{n}: \tilde{n} \neq n, \tilde{n} \in \text{bag}, H(n - \tilde{n}) = 0) \]

\#clashes \geq 1

1(\#clashes \geq 1)
\[ \leq \sum_{n} p(n) \sum_{\tilde{n} \neq n} 1[H(n-\tilde{n})=0] \]

- \( \tilde{n} \neq n \)
- \( \tilde{n} \in \text{bag} \)
\[ \sum_{n}^{\infty} P(n) \sum_{\tilde{n} \neq n, \tilde{n} \in \text{bag}} 1 \left[ H(n - \tilde{n}) = 0 \right] \]

\[ \langle P_{\Pi} \rangle = \sum_{H} P(H) \sum_{n \in \text{bag}} P(n) \sum_{\tilde{n} \in \text{bag}, \tilde{n} \neq n} 1 \left( H_x = 0 \right) \]

where \( x = n - \tilde{n} \)
If \( \mathbf{x} \) is a fixed non-zero binary vector of length \( N \) and \( \mathbf{h} \) is a random \((\frac{1}{2}, \frac{1}{2})\) binary vector of length \( N \), what is the probability that

\[ \mathbf{h}^\top \mathbf{x} = 0 \mod 2? \]

\[ e.g. \quad \mathbf{x} : \quad 0 \quad 1 \quad 1 \quad 1 \quad 1 \]
\[ \mathbf{h} : \quad h_1 \quad h_2 \quad h_3 \quad h_4 \]
What is the probability that \( h^\pi \cdot x = 0 \)?

Example: \( h \cdot x = h_2 + h_3 + h_4 \mod 2 \)
If $\mathbf{x}$ is a fixed non-zero binary vector of length $N$ and $\mathbf{H}$ is a random $(\frac{1}{2}, \frac{1}{2})$ binary matrix,

$$
\mathbf{H} = \begin{bmatrix}
\leftarrow & h^{(1)} & \rightarrow \\
\leftarrow & h^{(2)} & \rightarrow \\
\vdots & \vdots & \vdots \\
\leftarrow & N & \rightarrow
\end{bmatrix}
$$

what is the probability that

$$\mathbf{H}\mathbf{x} = 0 \mod 2?$$
What is Prob that $h \cdot x = 0$?

$\frac{1}{2}$

eg $h \cdot x = h_2 + h_3 + h_4 \mod 2$

What is Prob $(h \cdot x = 0)$?

$\sum p(n) \sum \left(\frac{1}{2}\right)^M \leq 1 \times 2 \text{NH}_2(\text{f}^*) \left(\frac{1}{2}\right)^M$
\[ \frac{1}{2} \text{ error} \leq P_{\text{err}} + \frac{1}{2} \frac{M - NH_2(f)}{2} \]

vanishes if \( M \gg NH_2(f) \)

ie \[ \frac{M}{N} > H_2(f) \]
\[ 1 - \frac{M}{N} < 1 - H_2(f) \]
\[ 1 - \frac{M}{N} \]

\[ \text{if } R < C \]

\[ Z = H(t + n) = H_n = H_\alpha \]

\[ QED \]

\[ n \rightarrow 000010100 \ldots \ldots \quad 000 \]

\[ #15 = NF + \alpha \sqrt{NF} \]

\[ t = \overline{\underline{\underline{\underline{\underline{\underline{SSS}}}}} \quad \underline{\underline{\underline{Sk)}}}} \quad t_n \]

\[ K \quad M \]
Homework recommendations

- Noisy channels - Chapters 8, 9, 10 (10.1-10.4 only)
  - Exercises 9.17 (p155); 10.12 (172); 15.12 (235)
  - and (if you want more practice) 15.11, 15.13, 15.15

- Invent a channel to pose to your colleagues:
  - 'what's the capacity of _this_?'