Neural dynamics of vision
A computational perspective

Clayton Seitz[1]

January 26, 2021

[1] cwseitz.github.io

Dedicated to Calvin and Hobbes.

Contents

1 The Neural Code **3**
1.1 Section heading 3

2 Learning Theory **5**
2.1 Section heading 5

3 Biologically-Inspired Computer Vision **7**
3.1 Natural Image Statistics 7
3.2 Gabor Analysis 7

4 Semantic Coding **9**
4.1 Section heading 9

5 Information and Coding Theory **11**
5.1 Introduction 11
5.2 Entropy 11
5.3 Joint Entropy 12
5.4 Source Coding 12
5.4.1 Kraft's Inequality 13
5.4.2 Jensen's Inequality 13
5.4.3 The Data-Processing Inequality 14
5.4.4 Example 1: Applying Jensen's Inequality 14
5.4.5 Example 2: Proving Cauchy-Schwarz 14
5.5 Error Correcting Codes 15

6 Microscopy and Image Analysis **17**
6.1 Section heading 17

Preface

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis risus ante, auctor et pulvinar non, posuere ac lacus. Praesent egestas nisi id metus rhoncus ac lobortis sem hendrerit. Etiam et sapien eget lectus interdum posuere sit amet ac urna.

Un-numbered sample section

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis risus ante, auctor et pulvinar non, posuere ac lacus. Praesent egestas nisi id metus rhoncus ac lobortis sem hendrerit. Etiam et sapien eget lectus interdum posuere sit amet ac urna. Aliquam pellentesque imperdiet erat, eget consectetur felis malesuada quis. Pellentesque sollicitudin, odio sed dapibus eleifend, magna sem luctus turpis.

Another sample section

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis risus ante, auctor et pulvinar non, posuere ac lacus. Praesent egestas nisi id metus rhoncus ac lobortis sem hendrerit. Etiam et sapien eget lectus interdum posuere sit amet ac urna. Aliquam pellentesque imperdiet erat, eget consectetur felis malesuada quis. Pellentesque sollicitudin, odio sed dapibus eleifend, magna sem luctus turpis, id aliquam felis dolor eu diam. Etiam ullamcorper, nunc a accumsan adipiscing, turpis odio bibendum erat, id convallis magna eros nec metus.

Structure of book

Each unit will focus on <SOMETHING>.

About the companion website

The website[1] for this file contains:

- A link to (freely downlodable) latest version of this document.
- Link to download LaTeX source for this document.
- Miscellaneous material (e.g. suggested readings etc).

Acknowledgements

- A special word of thanks goes to Professor Don Knuth[2] (for TEX) and Leslie Lamport[3] (for LATEX).
- I'll also like to thank Gummi[4] developers and LaTeXila[5] development team for their awesome LATEX editors.
- I'm deeply indebted my parents, colleagues and friends for their support and encouragement.

Amber Jain
`http://amberj.devio.us/`

[1] `https://github.com/amberj/latex-book-template`
[2] `http://www-cs-faculty.stanford.edu/~uno/`
[3] `http://www.lamport.org/`
[4] `http://gummi.midnightcoding.org/`
[5] `http://projects.gnome.org/latexila/`

1

The Neural Code

"This is a quote and I don't know who said this."
– Author's name, *Source of this quote*

1.1 Section heading

2

Learning Theory

"This is a quote and I don't know who said this."
– Author's name, *Source of this quote*

2.1 Section heading

3

Biologically-Inspired Computer Vision

“This is a quote and I don’t know who said this.”
– Author’s name, *Source of this quote*

3.1 Natural Image Statistics

3.2 Gabor Analysis

4

Semantic Coding

"This is a quote and I don't know who said this."
– Author's name, *Source of this quote*

4.1 Section heading

5

Information and Coding Theory

> *"We may have knowledge of the past but cannot control it; we may control the future but have no knowledge of it"*
>
> – Claude Shannon

5.1 Introduction

Information theory is a framework first introduced by Claude Shannon's seminal paper *A mathematical theory of communication* published in 1948. At it's core, information theory makes the intuitive concept of *information* mathematically rigorous and forms the foundation of many modern communication systems. Neural circuits in the visual system are an especially interesting example of such a communication system. Therefore, in this section, the information theoretic concepts necessary for studying neural circuits are introduced.

5.2 Entropy

The concept of entropy is not exclusive to information theory; rather, it is used widely in disciplines such as physics and mathematical statistics. In fact, entropy was originally defined in statistical physics when Ludwig Boltzmann gave a statistical description of a thermodynamic system of particles. Since this is arguably the more intuitive path as opposed an entirely mathematical description, I will follow a similar line of reasoning in the following paragraphs.

In every application, the entropy $\mathbf{H}$ is a measure of uncertainty or how much information is contained in a random variable x. In information theory, the entropy is a property of a probability distribution of a random variable

$P(x)$ where x can take on continuous or discrete values. For the discrete case, we can express the entropy in bits

$$\mathbf{H} = \sum_{x \in S} P(x) \log \frac{1}{P(x)} \tag{5.1}$$

where the set S spans the entire space of possible discrete values of x. We can go on to derive upper and lower bounds for the entropy. Notice that $\mathbf{H} \geq 0$ since $P(x) \leq 1$ and therefore $\log P(x) \leq 0$ for all x. At the same time, if we define a variable $Y = \frac{1}{\log x}$, we can write

$$\begin{aligned}
\mathbf{H} &= \mathbf{E}[\log Y] \\
&\leq \log \mathbf{E}[Y] \\
&= \log \sum_{y} P(x) \frac{1}{P(x)} \\
&= \log |S|
\end{aligned}$$

which is just the entropy of a uniform distribution.

5.3 Joint Entropy

$$\begin{aligned}
\mathbf{H}(X, Y) &= \sum_{x,y} P(x, y) \log \frac{1}{P(x, y)} \\
&= \sum_{x,y} P(x) P(y|x) \log \frac{1}{P(x) P(y|x)} \\
&= \sum_{x,y} P(x) P(y|x) \log \frac{1}{P(x)} + \sum_{x,y} P(x) P(y|x) \log \frac{1}{P(y|x)} \\
&= \sum_{x,y} P(x) P(y|x) \log \frac{1}{P(x)} + \sum_{x} P(x) \sum_{y} P(y|x) \log \frac{1}{P(y|x)} \\
&= H(X) + H(Y|X)
\end{aligned}$$

This result defines the **chain rule** for entropy.

5.4 Source Coding

Definition 1. *A code of a set S that uses an alphabet Ω is a map $C : S \to \Omega$ that assigns each element of S a finite string over the alphabet Ω. We say that the mapping C is* **prefix free** *if for all pairs $x, y \in S$ where $x \neq y$, $C(x)$ is not a prefix of $C(y)$.*

Most of the time the alphabet Ω we use is the set $0, 1$.

5.4.1 Kraft's Inequality

Definition 2. *For a binary code, there exists a prefix free code C with codeword lengths l_i if and only if*

$$\sum_i 2^{-l_i} \leq 1 \tag{5.2}$$

At this point we would like to apply the concept of entropy to source coding. Indeed, it is true that if we have a random variable X over the set S, the minimum number of bits it will take us to communicate the value of X on average is the entropy $H(X)$.

Proof. The expected number of bits to communicate X is given by $\sum_x p(x)|C(x)|$

$$\begin{aligned} H(X) - \sum_x P(x)|C(x)| &= \sum P(x)[\log \frac{1}{P(x)} - |C(x)|] \\ &= \sum P(x) \log \frac{1}{P(x)2^{|C(x)|}} \\ &\geq \log \sum P(x) \frac{1}{P(x)2^{|C(x)|}} \\ &= \log \sum \frac{1}{2^{|C(x)|}} \\ &\leq 0 \end{aligned}$$

□

by Kraft's inequality for prefix-free codes.

5.4.2 Jensen's Inequality

Jensen's inequality is a statement about convexity. Consider a binary variable x that takes the value 0 with probability α and value 1 with probability $1 - \alpha$.

$$x = \begin{cases} 0 & \alpha \\ 1 & 1 - \alpha \end{cases}$$

A function f of the variable x is said to be *convex* if the following inequality holds

$$\alpha f(x) + (1 - \alpha) f(y) \leq f(\alpha x + (1 - \alpha) y)$$

which when generalized for an arbitrary random variable x forms Jensen's inequality

$$\mathbf{E}[f(x)] \leq f(\mathbf{E}[x]) \tag{5.3}$$

5.4.3 The Data-Processing Inequality

The data-processing inequality states that for any function f s.t. $y = f(X)$, we have that $H(Y) \geq H(X)$. In plain english, that means that the process of transforming X can never decrease it's entropy. Recall from our analysis of joint entropy above that

$$\begin{aligned} H(X,Y) &= H(X) + H(Y|X) \quad (5.4) \\ &= H(Y) + H(X|Y) \quad (5.5) \end{aligned}$$

If we can show that $H(y|x) = 0$ then the data-processing inequality holds. This is more obvious if you notice that if $y = f(x)$ then the distribution $P(x, y) = P(x)$ or $P(y|x) = 0$.

In terms of our venn diagram above, $H(y)$ is a *subset* of $H(x)$ which also implies the following for the mutual information: $I(x, y) = H(y)$.

5.4.4 Example 1: Applying Jensen's Inequality

Let's consider a function $f : \mathbb{R} \to \mathbb{R}$. Using Jensen's inequality, we can prove that $f = x^2$ or $f = x \log x$ are convex functions. Let's begin by applying it to x^2 for a general normalized probability distribution $p(x)$.

$$\begin{aligned} \int p(x)f(x)dx &= \int x^2 p(x)dx \\ &= x^2 - 2\int xdx \\ &= 0 \leq x^2 \ \forall x \end{aligned}$$

We have a similar proof for $f(x) = x \log x$

$$\begin{aligned} \int p(x)f(x)dx &= \int x \log x \ p(x)dx \\ &= x \log x - \int \frac{d}{dx} x \log x \ dx \\ &= 0 \leq \mu \log \mu \end{aligned}$$

where $\mu = \mathbf{E}[x] \geq 0$ since f is only defined on $[0, \infty]$.

5.4.5 Example 2: Proving Cauchy-Schwarz

A common form of the Cauchy-Schwarz inequality states that for two vectors u and v, we have

$$u \cdot v \leq \|u\| \, \|v\|$$

5.5 Error Correcting Codes

6

Microscopy and Image Analysis

"This is a quote and I don't know who said this."
– Author's name, *Source of this quote*

6.1 Section heading