

# Exercise sheet 1

March 18, 2026

The following are exercises on natural latents. Each problem is broken into a sequence of lemmas leading to a main theorem. **For each subquestion, try to prove the stated lemma before reading on.** If you get stuck, you may treat the lemma as given and proceed to the next part.

## Natural latent Setup

Let  $X_1, \dots, X_n$  be observable random variables and  $\Lambda$  a latent variable. We write  $X = (X_1, \dots, X_n)$ . All entropy and mutual information quantities are understood with respect to the joint distribution  $P[X, \Lambda]$ .

### Key definitions.

- **Mediation.**  $\Lambda$  is an exact *mediator* over  $X_1, \dots, X_n$  if

$$P[X_1, \dots, X_n | \Lambda] = \prod_i P[X_i | \Lambda],$$

i.e.  $X_1, \dots, X_n$  are conditionally independent given  $\Lambda$ . The *mediation error* of a latent variable  $\Lambda$  is

$$\varepsilon_{\text{med}}(\Lambda) = D_{\text{KL}}(P[X_1, \dots, X_n, \Lambda] \| P[\Lambda] \prod_i P[X_i | \Lambda]).$$

- **Redundancy.**  $\Lambda$  is an exact *redund* over  $X_1, \dots, X_n$  if  $\Lambda$  is fully determined by each  $X_i$  individually, i.e.  $H(\Lambda | X_i) = 0$  for all  $i$ . The *redundancy error* of a latent variable  $\Lambda$  w.r.t  $X_i$  is

$$\varepsilon_{\text{red}}(\Lambda) = \sum_i H(\Lambda | X_i).$$

- **Naturality.**  $\Lambda$  is a *natural latent* over  $X_1, \dots, X_n$  if it satisfies both mediation and redundancy (exactly, or approximately up to some  $\varepsilon > 0$ ).

**Notation.** We write  $TC(X | \Lambda)$  for the *conditional total correlation*:

$$TC(X | \Lambda) = \sum_i H(X_i | \Lambda) - H(X | \Lambda).$$

## Exercise 1: Mediation error as conditional total correlation

Throughout this exercise, assume  $\Lambda$  is a **deterministic function of  $X$** , i.e.  $H(\Lambda | X) = 0$ .

**Part 1(a).** Show that the mediation error equals the conditional total correlation:

$$\varepsilon_{\text{med}}(\Lambda) = TC(X | \Lambda) = \sum_i H(X_i | \Lambda) - H(X | \Lambda).$$

You may find the following identity useful:

- **Chain rule for KL divergence.** For random variables  $A, B$ :

$$D_{\text{KL}}(P[A, B] \| Q[A, B]) = D_{\text{KL}}(P[A] \| Q[A]) + \mathbb{E}_A[D_{\text{KL}}(P[B | A] \| Q[B | A])].$$

**Part 1(b).** Show that if  $\Lambda$  is a deterministic function of  $X$ , then

$$H(X | \Lambda) = H(X) - H(\Lambda).$$

You may find the following identity useful:

- **Mutual information.**  $I(X; Y) = H(X) - H(X | Y)$

**Part 1(c).**

Let  $\mathcal{A}_k = \{\Lambda' : \exists f \text{ s.t. } \Lambda = f(X), H(\Lambda') = k\}$  be the set of latent variables that are deterministic functions of  $X$  with entropy exactly  $k$ .

Show that  $\Lambda \in \mathcal{A}_k$  minimizes the mediation error  $\varepsilon_{\text{med}}(\Lambda)$  (among latent variables in  $\mathcal{A}_k$ ) if and only if it maximizes  $\sum_i I(X_i; \Lambda)$ .

## Exercise 2

**Part 2(a).**

Let  $\mathcal{A}_k = \{\Lambda' : H(\Lambda' | X) = 0, H(\Lambda') = k\}$  as before.

Show that  $\Lambda \in \mathcal{A}_k$  minimizes the redundancy error

$$\varepsilon_{\text{red}}(\Lambda) = \sum_i H(\Lambda | X_i)$$

if and only if it maximizes  $\sum_i I(X_i; \Lambda)$ .

**Part 2(b).**

Define

$$\varepsilon_{\text{med}}^*(k) = \min_{\Lambda' \in \mathcal{A}_k} \varepsilon_{\text{med}}(\Lambda'), \quad \varepsilon_{\text{red}}^*(k) = \min_{\Lambda' \in \mathcal{A}_k} \sum_i \varepsilon_{\text{red}_i}(\Lambda').$$

Assume that  $\varepsilon_{\text{red}}^*(k)$  weakly increases with  $k$  and  $\varepsilon_{\text{med}}^*(k)$  weakly decreases with  $k$ .

Define the (weak) *Pareto frontier of naturality* of  $X$  as the set of latent variables that are (weakly) Pareto-optimal with respect to the mediation and redundancy errors  
i.e. no other latent variable achieves a strictly lower error on both errors.

Characterize the Pareto frontier of naturality. What are the parameters that matter?

*Hint: Make use of Parts 1(c) and 2(a), and the provided assumption*

## Exercise 3

### Definitions.

We distinguish two related but distinct notions for mediators and redunds:

- An *minimal mediator* over  $X_1, \dots, X_n$  is a mediator  $\Lambda$  such that for any other mediator  $\Lambda'$ , we have  $H(\Lambda | \Lambda') < \varepsilon$ , i.e.  $\Lambda$  is an approximate deterministic function of every other mediator.
- A *minimum mediator* over  $X_1, \dots, X_n$  is an exact mediator  $\Lambda^*$  with the smallest entropy  $H(\Lambda^*)$  among all mediators.
- An *maximal redund* over  $X_1, \dots, X_n$  is a redund  $\Lambda$  such that for any other redund  $\Lambda'$ , we have  $H(\Lambda' | \Lambda) < \varepsilon$ , i.e. every other redund is an approximate deterministic function of  $\Lambda$ .
- A *maximum redund* over  $X_1, \dots, X_n$  is an exact redund  $\Lambda_*$  with the largest entropy  $H(\Lambda_*)$  among all redunds.

Conceptually, a minimal mediator need not be a minimum mediator, and a maximal redund need not be a maximum redund. The exercises below establish connection between the two.

### Part 3(a).

Let  $\Lambda$  and  $\Lambda'$  be two minimal mediators over  $X_1, \dots, X_n$ . Show that they are approximately isomorphic, i.e.

$$H(\Lambda | \Lambda') < \varepsilon \quad \text{and} \quad H(\Lambda' | \Lambda) < \varepsilon.$$

Conclude that any two maximal redunds are likewise approximately isomorphic.

### Part 3(b).

Let  $\Lambda$  be a minimal mediator and  $\Lambda^*$  a minimum mediator over  $X_1, \dots, X_n$ . Show that they are approximately isomorphic up to a constant multiple of  $\varepsilon$ , i.e. there exists a small constant  $c$  such that

$$H(\Lambda | \Lambda^*) < c\varepsilon \quad \text{and} \quad H(\Lambda^* | \Lambda) < c\varepsilon.$$

Prove the analogous result for any maximal redund  $\Lambda$  and maximum redund  $\Lambda_*$ : they are approximately isomorphic up to a constant multiple of  $\varepsilon$ .

### Part 3(c).

Recall from Exercise 2 the assumption that  $\varepsilon_{\text{red}}^*(k)$  weakly increases with  $k$  and  $\varepsilon_{\text{med}}^*(k)$  weakly decreases with  $k$ , where

$$\varepsilon_{\text{med}}^*(k) = \min_{\Lambda' \in \mathcal{A}_k} \varepsilon_{\text{med}}(\Lambda'), \quad \varepsilon_{\text{red}}^*(k) = \min_{\Lambda' \in \mathcal{A}_k} \varepsilon_{\text{red}}(\Lambda').$$

We say that moving from entropy level  $k_1$  to  $k_2$  is a *Pareto-improvement on naturality* if

$$\varepsilon_{\text{med}}^*(k_2) \leq \varepsilon_{\text{med}}^*(k_1) \quad \text{and} \quad \varepsilon_{\text{red}}^*(k_2) \leq \varepsilon_{\text{red}}^*(k_1),$$

with at least one inequality strict. In other words, moving from  $k_1$  to  $k_2$  does not worsen either error and strictly improves at least one.

Using the Pareto frontier characterised in Exercise 2(b), answer the following:

- For which values of  $k$  is a Pareto-improvement on naturality possible by increasing  $k$ ? By decreasing  $k$ ?

- For which values of  $k$  is there a real tradeoff between the mediation and redundancy errors, i.e. any change in  $k$  strictly worsens at least one error?

*Hint: Consider the definitions of the maximum redund  $\Lambda_*$  and minimum mediator  $\Lambda^*$ , and what they imply about  $\varepsilon_{\text{red}}^*(k)$  and  $\varepsilon_{\text{med}}^*(k)$  at  $k = H(\Lambda_*)$  and  $k = H(\Lambda^*)$  respectively.*

**Part 3(d).**

A *natural latent* over  $X_1, \dots, X_n$  is a latent variable  $\Lambda$  that is simultaneously a maximal redund and a minimal mediator.

Using the approximate isomorphism between maximal and maximum redunds (and between minimal and minimum mediators) established in Parts 3(a) and 3(b), show that if a natural latent exists for  $X$ , then  $H(\Lambda_*) \approx H(\Lambda^*)$ .

What does the existence of natural latent for  $X$  say about the Pareto frontier of naturality of  $X$ ?