

Exercise sheet 2

March 18, 2026

*The following are exercises on condensation and 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.*

Setup

Let X_1, \dots, X_n be observable random variables with index set $I = \{1, \dots, n\}$. We write \mathcal{P}^+I for the set of all non-empty subsets of I .

Latent Variable Model (LVM).

A latent variable model over X_1, \dots, X_n is a collection of latent variables $(Y_A)_{A \in \mathcal{P}^+I}$, where Y_A is indexed by a non-empty subset $A \subseteq I$. The index A indicates which observables Y_A contributes to. Latent variables that are constant (carrying no information) are omitted. We require the *latent variable model condition*: each X_i is exactly recoverable from $Y_{\supset i} = (Y_A : i \in A)$, the collection of all latent variables above i .

Condensation.

For a latent variable model $(Y_A)_{A \in \mathcal{P}^+I}$, define the *conditioned score* for a non-empty subset $A \subseteq I$ as

$$\chi(A) = \sum_{B \cap A \neq \emptyset} H(Y_B | Y_{\supset B}),$$

where $Y_{\supset B} = (Y_C : C \supsetneq B)$ denotes the latent variables strictly above B . The conditioned score satisfies $\chi(A) \geq H(X_A)$ always.

We say that $(Y_A)_{A \in \mathcal{P}^+I}$ is an ε -*approximate condensation* of X_1, \dots, X_n if

$$\chi(A) \leq H(X_A) + \varepsilon, \quad \forall A \in \mathcal{P}^+I.$$

A *perfect condensation* is a 0-approximate condensation, i.e. $\chi(A) = H(X_A)$ for all A .

Strong and Weak Natural Latents.

Recall from Exercise Sheet 1 that an ε -*approximate (strong) natural latent* Λ over X_1, \dots, X_n satisfies:

- **Mediation:** $\text{Med}(\Lambda) = \sum_{i=1}^n H(X_i | \Lambda) - H(X_1, \dots, X_n | \Lambda) \leq \varepsilon$
- **Strong redundancy:** $H(\Lambda | X_i) \leq \varepsilon$ for all $i \in I$

We say Λ is an ε -*approximate weak natural latent* over X_1, \dots, X_n if it satisfies mediation and the following weaker redundancy condition:

- **Weak redundancy:** $H(\Lambda | X_{I \setminus \{i\}}) \leq \varepsilon$ for all $i \in I$,

where $X_{I \setminus \{i\}} = (X_j : j \neq i)$ denotes all observables except X_i . That is, Λ can be recovered from all-but-one of the observables, rather than from any single one.

Exercise 1

Part 1(a).

Suppose we have an ε -approximate condensation $(Y_A)_{A \in \mathcal{P}+I}$ of X_1, \dots, X_n . Fix a non-empty subset $A \subseteq I$. We construct a candidate weak natural latent over $(X_i)_{i \in A}$ as follows:

- Define $\Lambda = (Y_B : B \cap A \neq \emptyset, |B \cap A| \geq 2)$, i.e. the collection of all latent variables that contribute to at least two observables in A .
- For each $i \in A$, define $Y'_{\{i\}} = (Y_B : B \cap A = \{i\})$, i.e. the collection of all latent variables whose only overlap with A is the single index i .

Show that for each $i \in A$, X_i is a deterministic function of $(\Lambda, Y'_{\{i\}})$, and use this to conclude that

$$H(X_i | \Lambda) \leq H(Y'_{\{i\}} | \Lambda). \quad (1)$$

Hint: Use the latent variable model condition, and consider what $Y_{\ni i}$ looks like in relation to Λ and $Y'_{\{i\}}$

Part 1(b).

Show that

$$\sum_{i \in A} H(Y'_{\{i\}} | \Lambda) - \varepsilon \leq H(X_A) - H(\Lambda) \leq H(X_A | \Lambda). \quad (2)$$

Hint: For the leftmost inequality, proceed in two steps:

1. Show that $\sum_{i \in A} H(Y'_{\{i\}} | \Lambda) + H(\Lambda) \leq \chi_L(A)$ by expanding $H(\Lambda)$ and each $H(Y'_{\{i\}} | \Lambda)$ via the chain rule. In each case, use the fact that conditioning can only reduce entropy.
2. Apply the definition of ε -approximate condensation to bound $\chi_L(A)$, then rearrange.

Part 1(c).

Using (1) and (2), show that the mediation error of Λ is bounded by ε :

$$\text{Med}(\Lambda) = \sum_{i \in A} H(X_i | \Lambda) - H(X_A | \Lambda) \leq \varepsilon.$$

Exercise 2

Part 2(a).

For any non-empty subset $B \subseteq I$, let $Y_{\cap B} = (Y_C : C \subseteq B, C \neq \emptyset)$ denote the collection of all latent variables whose index set is contained within B . Assume the following bound (Proposition 5.2 in the condensation paper):

$$H(Y_{\cap B}) \leq \chi_L(B).$$

Using this and the definition of ε -approximate condensation, show that

$$H(Y_{\cap B} | X_B) \leq \varepsilon.$$

Hint: Use the fact that X_B is a deterministic function of $Y_{\cap B}$ (by the latent variable model condition).

Part 2(b).

Using Part 2(a), show that Λ satisfies weak redundancy, i.e.

$$H(\Lambda | X_{A \setminus \{i\}}) \leq \varepsilon, \quad \forall i \in A.$$

Hint: Set $B = A \setminus \{i\}$ in Part 2(a), and consider the relationship between Λ and $Y_{\cap(A \setminus \{i\})}$.

Remark.

Combining the mediation bound from Exercise 1(c) with the weak redundancy bound from Part 2(b), we have shown that the construction $\Lambda = (Y_B : B \cap A \neq \emptyset, |B \cap A| \geq 2)$ is an ε -approximate weak natural latent over $(X_i)_{i \in A}$. That is, any ε -approximate condensation of X_1, \dots, X_n gives rise to an ε -approximate weak natural latent over any subset of the observables.

Exercise 3

Setup.

Given an ε -approximate condensation $(Y_B)_{B \in \mathcal{P}+I}$ satisfying $\chi_L(B) \leq H(X_B) + \varepsilon$ for all B , fix a subset $A \subseteq I$ and define:

- $\Lambda = Y_{\supseteq A} = (Y_B : B \supseteq A)$, the collection of all latent variables whose index set contains A .
- $\Lambda' = (Y_B : B \cap A \neq \emptyset, |B| \geq 2, B \not\supseteq A)$, the collection of non-singleton latent variables that partially overlap with A but do not contain it.
- $\Lambda^* = (\Lambda, \Lambda')$, the combined latent variable.

Assume additionally that $H(\Lambda' | \Lambda) \leq \varepsilon$, meaning there is approximately no lower-level structure among the variables in A beyond what is captured by Λ .

The goal of this exercise is to show that Λ is an $(n + 1)\varepsilon$ -approximate strong natural latent over $(X_i)_{i \in A}$.

Strong Redundancy

Part 3(a).

Show that Λ satisfies strong redundancy:

$$H(\Lambda | X_i) \leq \varepsilon, \quad \forall i \in A.$$

Hint: Recall Part 2(b).

Mediation

Part 3(b).

Show that for any random variable Z ,

$$H(Z | \Lambda) \leq H(Z | \Lambda^*) + H(\Lambda^* | \Lambda). \tag{1}$$

Hint: Write $H(Z, \Lambda) \leq H(Z, \Lambda, \Lambda^)$, expand using chain rule.*

Part 3(c).

Argue that

$$H(X_A | \Lambda^*) \leq H(X_A | \Lambda). \tag{2}$$

Part 3(d).

Using (1) and (2) together with the assumption $H(\Lambda' | \Lambda) \leq \varepsilon$ and the fact that Λ^* is an ε -approximate weak natural latent (from Exercise 2), show that Λ satisfies mediation:

$$\text{Med}(\Lambda) = \sum_{i \in A} H(X_i | \Lambda) - H(X_A | \Lambda) \leq (n + 1)\varepsilon.$$

Exercise 4

Setup.

Given an ε -approximate natural latent Λ over $(X_i)_{i \in A}$, we construct a latent variable model $(Y_B)_{B \in \mathcal{P}^+ A}$ by setting:

- $Y_A = \Lambda$
- $Y_{\{i\}} = X_i$ for each $i \in A$
- All other latent variables empty (constant)

The goal of this exercise is to show that this construction yields a 2ε -approximate condensation over $(X_i)_{i \in A}$.

Part 4(a).

Show that the conditioned score satisfies

$$\chi(A) \leq H(X_A, \Lambda) + \varepsilon.$$

Hint: Use the mediation property of Λ

Part 4(b).

Using Part 4(a), show that the construction is a 2ε -approximate condensation, i.e.

$$\chi(A) \leq H(X_A) + 2\varepsilon,$$

Hint: Use the strong redundancy property of Λ .

Remark.

Both the mediation and strong redundancy properties of Λ transfer to any subset $B \subset A$. Therefore, the same argument applies with A replaced by B throughout, and we obtain the same 2ε bound on the conditioned score.

Exercise 5 (More difficult)

Setup.

A *hierarchical natural latent* model formalises the idea that knowledge can be organised into a nested hierarchy of concepts: at the base level we have observables, at the next level we have natural latents that summarise shared information within small groups of observables, and at higher levels we have natural latents that summarise shared information across those groups, and so on. For example, individual images of cats and dogs are the observables; the features shared within each species (“cat-features”, “dog-features”) form first-order latents; and the features shared across all animals form a second-order latent.

Formally, we define a sequence of partitions Q^0, Q^1, \dots, Q^n of I , where:

- $Q^0 = \{\{i\} : i \in I\}$ is the finest partition (singletons).
- $Q^n = \{I\}$ is the coarsest partition (all of I).
- Q^{k-1} is a refinement of Q^k for each k : every block of Q^{k-1} is contained in a block of Q^k .

We then define

$$R = \{\Lambda_A^k : A \in Q^k, k \in 0 \dots n\},$$

where $\Lambda_{\{i\}}^0 = X_i$ are the observables, Λ_A^1 is an ε -approximate natural latent over $(X_i)_{i \in A}$ for each $A \in Q^1$, and Λ_A^k is an ε -approximate natural latent over $(\Lambda_B^{k-1})_{B \subset A, B \in Q^{k-1}}$ for each $A \in Q^k$, $k \geq 1$.

For $B \notin Q^k$ but $B \in Q^r$ for some $r > k$, we define $\Lambda_B^k = (\Lambda_C^k : C \subset B, C \in Q^k)$.

We further require the conditional independence condition: for each k and each $B \in Q^k$, Λ_B^{k-1} is conditionally independent of $\Lambda_{\bar{B}}^{k-1}$ given Λ_B^k (where \bar{B} denotes the complement of B within the next block up).

To construct an LVM from R , set $Y_A = \Lambda_A^k$ for all $\Lambda_A^k \in R$, and leave all other latent variables empty. The goal of this exercise is to show that this LVM is an approximate condensation.

Part 5(a).

Prove the following lemma: for each level k ,

$$\sum_{B \in Q^k} \sum_{C \subset B} H(\Lambda_C^{k-1} | \Lambda_B^k) \leq H(\Lambda_I^{k-1} | \Lambda_I^k) + K\varepsilon,$$

where K is a constant depending on the partition structure.

Hint: Use the conditional independence condition to show that $H(\Lambda_B^{k-1} | \Lambda_B^k) = H(\Lambda_B^{k-1} | \Lambda_I^k)$. Use the mediation property of Λ_B^k over $(\Lambda_C^{k-1})_{C \subset B}$ to bound $\sum_{C \subset B} H(\Lambda_C^{k-1} | \Lambda_B^k)$; remember that Q^k forms a partition.

Remark. For any subset $A \subseteq I$, an analogous version of the lemma holds with I replaced by $\cap A$ (restricting all variables and partitions to those that contribute to A):

$$\sum_{B \in Q^k, B \cap A \neq \emptyset} \sum_{C \subset B, C \cap A \neq \emptyset} H(\Lambda_C^{k-1} | \Lambda_B^k) \leq H(\Lambda_{\cap A}^{k-1} | \Lambda_{\cap A}^k) + K\varepsilon.$$

This holds because mediation transfers to subsets.

Part 5(b).

Using Part 5(a) and its subset version, show that for any $A \subseteq I$, the conditioned score satisfies

$$\chi(A) \leq H(\Lambda_{\cap A}^k, k = 0 \dots n) + K_2\varepsilon.$$

Hint: Note that $\Lambda_{\cap A}^k$ is an approximate deterministic function of $\Lambda_{\cap A}^{k-1}$ which means $H(Z | \Lambda_{\cap A}^k, \Lambda_{\cap A}^{k-1}) \approx H(Z | \Lambda_{\cap A}^{k-1})$

Part 5(c).

Using Part 5(b), show that

$$\chi(A) \leq H(X_A) + K_3\varepsilon,$$

concluding that the LVM constructed from R is a $K_3\varepsilon$ -approximate condensation.