STA414: Lecture 8 Quiz (Bayesian Methods in High Dimensions)

Instructor: Thibault Randrianarisoa

Questions

Question 1 What is a primary computational advantage of using a continuous shrinkage prior, such as the Horseshoe prior, over a traditional hard spike-and-slab prior in high-dimensional settings?

a) It avoids the need to explore a massive combinatorial space of possible models via MCMC sampling.

b) It yields a posterior that is asymptotically the exact product of normal distributions of the oracle dimension and Dirac masses at zero.

c) It restricts the prior to a single distribution that optimizes only the slab component without relying on a peak at zero.

d) It achieves exact sparsity by directly setting non-relevant parameters to precisely zero during the coordinate ascent updates.

Question 2 According to the lecture, what is a key theoretical limitation of using the full Bayesian posterior under a LASSO (Laplace) prior in sparse high-dimensional models?

a) The MAP estimator derived from the Laplace prior completely fails to induce sparsity.

b) The full posterior distribution does not contract at the same optimal speed as its mode.

c) The convergence rate of the LASSO mode is drastically slower than the optimal minimax rate.

d) It requires replacing the Laplace prior with a standard Normal prior to achieve any valid contraction.

Question 3 How does the Immersion Posterior (or projection-posterior) approach fundamentally differ from Variational Bayes (VB) when inducing sparsity?

a) Variational Bayes applies a projection map to the unconstrained coefficients, whereas Immersion Posterior uses Coordinate Ascent Variational Inference.

b) Immersion Posterior requires exploring an intractable combinatorial space of all possible models, whereas Variational Bayes strictly uses continuous shrinkage.

c) Immersion Posterior applies a sparsity-inducing optimization map to individual samples drawn from an unrestricted posterior, whereas VB performs optimization on distributions.

d) Variational Bayes is theoretically restricted to low-dimensional models, while Immersion Posterior is the only method capable of handling high dimensions.

Question 4 When utilizing mean-field spike-and-slab Variational Class ($\mathcal{Q}$) for high-dimensional inference with a spike-and-slab prior, what is a key theoretical property of this variational approach?

a) It requires the exact same slab distribution (e.g., Laplace) to be used within the variational class $\mathcal{Q}$ as is used in the prior.

b) It yields the same convergence rates and variable selection properties as the exact spike-and-slab posterior.

c) It mathematically forces all mixture weights γ_i to be strictly 0 or 1, discarding continuous uncertainty.

d) It suffers from a degraded convergence rate compared to the exact posterior, trading theoretical guarantees for computational speed.

Solutions

1. Correct Answer: a) It avoids the need to explore a massive combinatorial space of possible models via MCMC sampling.
Reasoning: Hard spike-and-slab priors require exploring 2^p possible models, which is computationally intractable in high dimensions, whereas continuous priors offer a more tractable 'soft' selection alternative.

2. Correct Answer: b) The full posterior distribution does not contract at the same optimal speed as its mode.
Reasoning: The Laplace prior faces conflicting demands for the penalty parameter (fitting zero parameters vs. non-zero ones), preventing the full posterior distribution from achieving the same optimal contraction rate as the LASSO point estimate.

3. Correct Answer: c) Immersion Posterior applies a sparsity-inducing optimization map to individual samples drawn from an unrestricted posterior, whereas VB performs optimization on distributions.
Reasoning: In the Immersion Posterior framework, optimization happens in the parameter space individually for each drawn sample, whereas VB performs a one-time optimization over families of distributions.

4. Correct Answer: b) It yields the same convergence rates and variable selection properties as the exact spike-and-slab posterior.
Reasoning: The theoretical properties of the mean-field variational class ensure that it matches the convergence rates and variable selection performance of the exact spike-and-slab posterior. Furthermore, option (a) is incorrect because Gaussian slabs are typically used within the variational class $\mathcal{Q}$ even when Laplace slabs are preferred in the prior. Option (c) is incorrect because γ_i takes continuous values in the interval $[0, 1]$.