Statistical Physics of Computation 2025 - Exercises

Vittorio Erba, Emanuele Troiani

Week 8

8.1 The average error of the BO estimator and of the Gibbs estimator

Consider a generic inference problem where you generate a hidden signal $X_* \sim P_0$, where P_0 is the prior distribution, and observe it through a noisy channel, obtaining the data/observation $Y \sim P_{\text{out}}$ ("out" stands for "output channel"). Think of X as a vector with N components, $X \in \mathbb{R}^N$, and Y as a vector with P components, $Y \in \mathbb{R}^P$. Consider the posterior distribution $P_{\text{posterior}}(X|Y) \propto P_0(X)P_{\text{out}}(Y|X)$, and suppose that you are in the Bayes Optimal (BO) setting, i.e. you know both P_0 and P_{out}, so you know the posterior.

We want to find expressions for the errors of the BO estimator w.r.t. to mean square error (MSE) loss (i.e. an expression for the MMSE) and of the Gibbs estimator, both as functions of the overlap order parameters associated to the posterior distribution. We define

$$
\begin{aligned}
Q_* &= \frac{1}{N}\mathbb{E}_{X\sim P_0}||X||^2 \\
Q &= \frac{1}{N}\mathbb{E}_Y\mathbb{E}_{X\sim P_{\text{posterior}}(\cdot|Y)}||X||^2 \\
m &= \frac{1}{N}\mathbb{E}_{Y,X_*}\mathbb{E}_{X\sim P_{\text{posterior}}(\cdot|Y)}X_*^T X \\
q &= \frac{1}{N}\mathbb{E}_Y\mathbb{E}_{X_1\sim P_{\text{posterior}}(\cdot|Y)}\mathbb{E}_{X_2\sim P_{\text{posterior}}(\cdot|Y)}X_1^T X_2
\end{aligned}
\tag{1}
$$

Q_* is the self-overlap (norm) of the signal, Q is the self-overlap (norm) of a sample from the posterior, m is the overlap of a sample from the posterior with the hidden signal, and q is the overlap between two independent samples from the posterior, and all of these quantities are averaged over the observation Y defining the posterior.

To be very explicit, the averages are defined as

$$
\begin{aligned}
\mathbb{E}_{X\sim P_0}f(X) &= \int dX\, f(X)P_0(X)\,, \\
\mathbb{E}_Y f(Y) &= \int dY\, dX_*\, f(Y)P_{\text{out}}(Y|X_*)P_0(X_*)\,, \\
\mathbb{E}_{Y,X_*}f(Y,X_*) &= \int dY\, dX_*\, f(Y,X_*)P_{\text{out}}(Y|X_*)P_0(X_*)\,, \\
\mathbb{E}_{X\sim P_{\text{posterior}}(\cdot|Y)}f(X) &= \int dX\, f(X)P_{\text{posterior}}(X|Y)\,.
\end{aligned}
\tag{2}
$$

These definitions (both the order parameters and the averages) should become natural to you, even if they are not right now. So convince yourself that they make sense, and learn them.

1. Show that for any estimator $\hat{X} : Y \to \hat{X}(Y)$, i.e. a function taking as input the observation, and outputting some estimate of the hidden signal that generated the observation, one has

$$\frac{1}{N}\mathbb{E}_{Y,X_*}||X_* - \hat{X}(Y)||^2 = \frac{1}{N}\mathbb{E}_{X^*\sim P_0}||X_*||^2 - \frac{2}{N}\mathbb{E}_{Y,X_*}X_*^T\hat{X}(Y) + \frac{1}{N}\mathbb{E}_Y||\hat{X}(Y)||^2 \,. \quad (3)$$

2. Show that $Q = Q_*$.

We start by considering the BO estimator. In class, we saw that the BO estimator w.r.t. the MSE is the average of the posterior, i.e.

$$\hat{X}_{\text{BO, MSE}}(Y) = \mathbb{E}_{X\sim P_{\text{posterior}}(\cdot|Y)}X \,. \quad (4)$$

3. Show that

$$\frac{1}{N}\mathbb{E}_{Y,X_*}||X_* - \hat{X}_{\text{BO, MSE}}(Y)||^2 = Q - 2m + q \,. \quad (5)$$

4. Argue finally that

$$\frac{1}{N}\mathbb{E}_{Y,X_*}||X_* - \hat{X}_{\text{BO, MSE}}(Y)||^2 = Q - q \,. \quad (6)$$

We now instead consider the Gibbs estimator. The Gibbs estimator estimates the signal by sampling from the posterior, instead of averaging the posterior. In this case, we consider the error of the Gibbs estimator on average, as it is a random estimator, where we mean average over the sampling procedure.

4. Show that

$$\frac{1}{N}\mathbb{E}_{Y,X_*}\mathbb{E}_{X\sim P_{\text{posterior}}(\cdot|Y)}||X_* - X||^2 = 2(Q - q) \,. \quad (7)$$

Notice the factor 2 difference with the BO estimator.

5. Argue that the Gibbs sampler is on average worse than the BO estimator.

6. When is the Gibbs sampler (on average) as effective as the BO estimator?

8.2 Bayesian learning of a scalar variable

We now consider the following scalar inference problem. We generate a hidden signal $x^* \in \mathbb{R}$ from a prior distribution $\mathrm{P}_0(x)$. We then observe only a noisy version of the signal

$$y = x^* + \sqrt{\Delta}z \quad (8)$$

where z is an independent Gaussian variable with mean zero and variance 1, and $\Delta > 0$ plays the role of a noise-to-signal ratio. Given y, we want to Bayes-optimally estimate the signal x^*. We first need to set up the Bayesian machinery.

1. Write the output channel distribution $P_{\text{out}}(y|x)$, i.e. the probability of observing y given a certain signal x.

2. Use Bayes theorem to show that the posterior distribution, i.e. the probability that the signal is x given our observation y, satisfies

$$P_{\text{posterior}}(x|y) = P_0(x)\frac{e^{-\frac{(y-x)^2}{2\Delta}}}{Z} \quad (9)$$

where Z is the normalization factor.

3. We need to find a good estimator $\hat{x}$ for our signal. We will use the Bayes Optimal estimator with respect to the MSE, i.e. the mean of the posterior distribution. Argue that we have:

$$\hat{x}(y) = \frac{\int \mathrm{d}x \, x \, P_0(x) e^{-\frac{(y-x)^2}{2\Delta}}}{\int \mathrm{d}x P_0(x) e^{-\frac{(y-x)^2}{2\Delta}}} \tag{10}$$

4. Suppose now that P_0 is a standard Gaussian. Show that

$$\hat{x}(y) = \frac{y}{1+\Delta} \, . \tag{11}$$

5. Show that the MSE that this estimator achieves is

$$\mathbb{E}_{x^*,y}\left[(\hat{x}(y) - x^*)^2\right] = \frac{\Delta}{1+\Delta} \tag{12}$$

6. If $\Delta \to \infty$ there is no information about the signal in the observation, as it is fully erased by the large amount of noise. In such a case, given that we know the prior, we may just sample a candidate estimate for the lost signal from the prior and hope that it achieves a good performance. What is, on average, the MSE of a sample from the prior? What is the MSE of the BO estimator when $\Delta \to \infty$? Discuss what is the BO estimator doing in this limit.

7. We now change the prior P_0. Find the BO estimator w.r.t. to the MSE error for the case in which $x^* = 1$ with probability p and zero otherwise. You should get:

$$\hat{x}(y) = \frac{1}{1 + \exp\left\{\frac{1-2y}{2\Delta}\right\} \frac{1-p}{p}} \tag{13}$$

Notice that in the last point we obtained an estimator which gives us as an estimate a continuous value in $[0,1]$, instead of just telling us whether the signal was $x_* = 1$ or $x_* = 0$. This is expected, as we asked for the estimate that minimizes the MSE, without specifying that it should be related to the support of the prior P_0. If we wanted an estimator respecting the constraint that $x \in \{0,1\}$, we could have used the maximum-a-posteriori estimator (MAP), i.e.

$$\hat{x}_{\text{MAP}}(y) = \text{argmax}_{x \in \{0,1\}} P_{\text{posterior}}(x|y) \tag{14}$$

giving as an estimate the most likely signal to have generated the observation y.

8. Compute the MAP estimator for the prior of point 7.