Lecture I: Denoising & Statistics

In a simple example:
A denoising key problem: vector components but obscure space signal

It turns out that denoising slowly the noise we go through a threshold value of the noise below which the denoising becomes easy after having been impossible.

Formalization: signal of length $n^2$

How components will have value $t \in [0, w, 0, w]$, on avg. $\sum x_i \in \frac{1}{\sqrt{n}} \sum w_i$, $w \approx \{ \Delta \rightarrow \infty \} \rightarrow \infty$ $\Rightarrow \frac{1}{\sqrt{2 \pi \Delta}} \exp[-\ln^2 \Delta / 2 \Delta]$

$\Rightarrow \frac{1}{\sqrt{2 \pi \Delta}} \exp[-\ln^2 \Delta / 2 \Delta]$

If $|\ln| > \sqrt{2 \Delta} \Rightarrow \# of such components $ \rightarrow 0$

$\frac{\Delta}{2 \ln w_n}$ seems to be matching the little experiment limit on detectability!

(Benorok Tolstien 1998: universal denoising threshold)

In denoising of a single variable
Consider a random variable $X$ distributed according to $P_x(x)$. We want to transmit it through a noisy channel + Send ground truth $X \rightarrow X + \mathcal{N}(0, 1)$

Received $Y = \sqrt{\lambda}X + Z$ $\Rightarrow \lambda = \text{signal to noise ration (SNR)}$

$Z \sim \mathcal{N}(0, 1)$

Base: receiving many measurements:
\[
\begin{align*}
Y_1 &= \sqrt{\lambda}X + Z_1 \\
Y_2 &= \sqrt{\lambda}X + Z_2 \\
\vdots \\
Y_N &= \sqrt{\lambda}X + Z_N
\end{align*}
\]

we implicitly define the likelihood = conditional probability: $P_{Y|X}(y|X) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(\sqrt{\lambda}x-y)^2}{2\lambda}}$

$P(Y|X) = \prod_{i=1}^{N} P_{Y|X}(y_i|X) = \frac{1}{(2\pi)^{N/2}} e^{-\frac{1}{2}\sum_{i=1}^{N} (\sqrt{\lambda}y_i - y_i)^2}$

The maximum likelihood estimator: $\hat{x}(y) = \arg\max_{x} P_{X|Y}(x|y) = \arg\max_{x} \log P_{X|Y}(x|y)$

In our case:
\[
\hat{x}(y) = \arg\max_{x} \left( -\frac{(\sqrt{\lambda}x-y)^2}{2\lambda} + \text{const} \right) = \frac{1}{N\lambda} \sum_{i=1}^{N} y_i \approx \bar{x} \quad \text{(exercise)}
\]

The maximum likelihood scheme always works when $N \rightarrow \infty$ (Laplace 1810, Fisher 1912)

Second case: receiving only one measurement:

we'll need Bayes theorem:
\[
P(B|A) = \frac{P(A|B)P(B)}{P(A)}
\]

Laplace 1815, Bayes, pure 1763...
The tricky part: most of this time

\[ \text{posterior distribution: } P(x|y) = \frac{P(y|x) P(x)}{\text{normalization constant}} \]

\[ \text{evidence: } Z(y) = \int dx \ e^{-E(x,y)} \]

with \[ P(x) = \frac{1}{Z(y)} \exp(-\log p(x) - \log p(x)) = e^{-E(x,y)} \]

New from the single measurement: the most probable value of \( \hat{x}_{\text{MAP}} = \arg \max_x P(x|y) \)

\[ \text{Least-likely, with decreasing temperature to find the ground state} \]

Nevertheless, there a a subtlety here:

\[ \text{minimum mean squared error: } \hat{\text{R}}_{\text{MSE}}(y) \text{ such that } (\hat{\text{R}}_{\text{MSE}} - x^*)^2 \text{ is minimum.} \]

\[ \text{Consider the posterior task: } R = \int P(x|y) (\hat{\text{R}}(x) - x) dx \text{ average value of squared error.} \]

\[ \text{Minimize it: } \frac{\partial R}{\partial R} = 0 \Rightarrow \hat{\text{R}}_{\text{MSE}} = \frac{\int dx \ x P(x|y)}{\int dx \ P(x|y)} \]

\[ \text{Minimize } R = \frac{1}{\text{trace distribution}} \]

\[ \text{Good estimator = magnetization } \hat{\text{R}}_{\text{MSE}} = \langle x \rangle \]

Re: we could consider other risks, other moments of the gap:

The best strategy: problem dependent \( \langle 1 - x^2 \rangle = \hat{\text{R}}_{\text{MSE}} \text{ median}. \)

Yet again, you might prefer to minimize the number of errors: \( 1 - S_p(x, y) \)

\[ \text{The associated risk if: } R = \int P(x|y) (1 - S_p(x, y)) dx \]

\[ \text{Bayes optimal error} \]

\[ \hat{x} = \arg \max_{\hat{x}} P(x|y) \]

\[ \text{BAYES OPTIMAL ERROR} \]

\[ \hat{x}_{\text{MAP}} = \arg \max_x P(x|y) \]

\[ \hat{x}_{\text{MAP}} = \arg \max_x P(x|y) \]

\[ \text{MAP} \]

Back to our specific problem: Redefining convolution by the partition function \( Z(y) \):

\[ P(x|y) = \frac{1}{Z(y)} \exp \left[ -\frac{\lambda^2}{2} + y_1 x \right] P(x) \]

\[ \text{The minimal mean squared error } \text{MSE} = \mathbb{E}_{y|x^*} \left[ (\hat{x}_{\text{MSE}} - x^*)^2 \right] \]

\[ \text{MSE} = q_0 + q_0 - 2 m \text{ with} \]

\[ q = \mathbb{E}_{y|x^*} \langle x \rangle \]

\[ m = \mathbb{E}_{y|x^*} \langle x^2 \rangle \]

\[ q_0 = \mathbb{E}_{y|x^*} \langle x^2 \rangle \]

\[ \hat{x}_{\text{MSE}} = \mathbb{E}_{y|x^*} \left[ \langle x^2 \rangle \right] \]

\[ \text{A interpretation in terms of replicas:} \]

\[ \text{OF 2 replicas } q = \mathbb{E}_{y|x^*} \left[ \langle x^{(1)} \rangle \langle x^{(2)} \rangle \right] \]

\[ \text{ii/ NISHIMORI RELATIONS: } \]

\[ \mathbb{E}_{y,x}\left[ \langle f(x,x^*) \rangle \right] = \mathbb{E}_{y} \left[ \langle f(x^*, x^{(2)}) \rangle \right] \]

\[ \text{Ex: } m = q \Rightarrow \text{MSE} = q_0 - m \]

\[ \text{Overlap: } s_{dy,dx} P(y,x) \int dx \ P(x|y) f(x,x^*) = \int dy \ P(y) \int dx \ dn^* P(x^*|y) P(x|y) f(x,x^*) \]

\[ = \int dy P(y) \int dx dn^* \cdots \]