MDL exercises, ninth handout

(due April 27th, 14:00)

Consider MDL model selection between

$$\mathcal{M}_0 = \{P_{0,\sigma} : \sigma > 0\} \text{ and } \mathcal{M}_1 = \{P_{\delta,\sigma} : \sigma > 0, \delta \in \mathbb{R}\}$$

where $P_{\delta,\sigma}$ is the distribution under which $X_1, X_2, \ldots, X_n$ are i.i.d., each with density given by

$$p_{\delta,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x}{\sigma} - \delta\right)^2}.$$

1. [**1 point**] Show that $\mathcal{M}_1$ is identical to the family of normal distributions with mean in $\mathbb{R}$ and variance in $\sigma^2 > 0$. That is, if $Q_{\mu,\sigma}$ represents a normal distribution with mean μ and variance σ, show that (i) for every $\sigma > 0, \delta \in \mathbb{R}$, there is a $\mu \in \mathbb{R}$ such that $P_{\delta,\sigma} = Q_{\mu,\sigma}$ and (ii), conversely, for every $\sigma > 0, \mu \in \mathbb{R}$, there is a $\delta \in \mathbb{R}$ such that $P_{\delta,\sigma} = Q_{\mu,\sigma}$.

   Solution: For every $\sigma > 0, \delta \in \mathbb{R}$, the density of $P_{\delta,\sigma}$ is given by

   $$\begin{aligned} p_{\delta,\sigma}(x) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x}{\sigma} - \delta\right)^2} \\ &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x - \delta\sigma}{\sigma}\right)^2} \\ &= q_{\sigma\delta,\sigma}, \end{aligned}$$

   where $q_{\sigma\delta,\sigma}$ is the density of the normal distribution $Q_{\sigma\delta,\sigma}$. Similarly we see that for every $\sigma > 0, \mu \in \mathbb{R}$: $Q_{\mu,\sigma} = P_{\mu/\sigma,\sigma}$.

We associate Bayesian universal measures $\bar{p}_0$ with $\mathcal{M}_0$ and $\bar{p}_1$ with $\mathcal{M}_1$. In both cases, we put the *right Haar prior* $\pi(\sigma) = 1/\sigma$ on the variance σ. For $\bar{p}_1$, we equip δ with some (arbitrary) proper prior density w. Thus, we measure the evidence against $\mathcal{M}_0$ by

$$M(x^n) := \log \frac{\bar{p}_1(x^n)}{\bar{p}_0(x^n)} \tag{1}$$

with $\bar{p}_0(x^n) = \int \sigma^{-1} p_{0,\sigma}(x^n) d\sigma$ and $\bar{p}_1(x^n) = \int_{\sigma>0,\delta\in\mathbb{R}} \sigma^{-1} w(\delta) p_{\delta,\sigma}(x^n) d\sigma d\delta$.

2. [**1 point**] Show that $\pi(\sigma) = 1/\sigma$ is improper.

Solution:

$$
\begin{aligned}
\int_0^\infty \pi(\sigma)d\sigma &= \int_0^\infty \frac{1}{\sigma}d\sigma \\
&= \lim_{\substack{a\to 0\\ b\to\infty}} \int_a^b \frac{1}{\sigma}d\sigma \\
&= \lim_{\substack{a\to 0\\ b\to\infty}} [\ln(\sigma)]_a^b \\
&= \lim_{\substack{a\to 0\\ b\to\infty}} (\ln(b)-\ln(a)) = \infty.
\end{aligned}
$$

3. (i) **[1 point]** Show that $M(x^n)$ is *scale-invariant.* That is, show that for every sequence $x_1,\ldots,x_n$, every $c>0$,

$$M(x_1,\ldots,x_n) = M(x_1/c,\ldots,x_n/c) \tag{2}$$

(HINT: re-express the integral over σ in $\bar{p}_0$ and $\bar{p}_1$ as an integral over $\sigma' = c\sigma$).

Solution:

For arbitrary x, we see

$$
\begin{aligned}
p_{\delta,\sigma}(x/c) &= \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x}{c\sigma}-\delta\right)^2} \\
&= \frac{c}{\sqrt{2\pi}c\sigma}e^{-\frac{1}{2}\left(\frac{x}{c\sigma}-\delta\right)^2} \\
&= cp_{\delta,c\sigma}(x).
\end{aligned}
$$

Then

$$p_{\delta,\sigma}(x_1/c,\ldots,x_n/c) = \prod_{i=1}^n p_{\delta,\sigma}(x_i/c) = c^n \prod_{i=1}^n p_{\delta,c\sigma}(x_i) = c^n p_{\delta,c\sigma}(x^n).$$

So we get

$$
\begin{aligned}
M(x^n/c) &= \log \frac{\overline{p}_1(x^n/c)}{\overline{p}_0(x^n/c)} \\
&= \log \frac{\int_{\sigma>0,\delta\in\mathbb{R}} \sigma^{-1}w(\delta)p_{\delta,\sigma}(x^n/c)d\sigma d\delta}{\int_{\sigma>0,\delta\in\mathbb{R}} \sigma^{-1}w(\delta)p_{0,\sigma}(x^n/c)d\sigma d\delta} \\
&= \log \frac{\int_{\sigma>0,\delta\in\mathbb{R}} \sigma^{-1}w(\delta)c^n p_{\delta,c\sigma}(x^n)d\sigma d\delta}{\int_{\sigma>0,\delta\in\mathbb{R}} \sigma^{-1}w(\delta)c^n p_{0,c\sigma}(x^n)d\sigma d\delta} \\
&= \log \frac{1/c\int_{\sigma>0,\delta\in\mathbb{R}} \sigma^{-1}w(\delta)p_{\delta,c\sigma}(x^n)d\sigma d\delta}{1/c\int_{\sigma>0,\delta\in\mathbb{R}} \sigma^{-1}w(\delta)p_{0,c\sigma}(x^n)d\sigma d\delta} \\
&= \log \frac{\int_{\sigma'>0,\delta\in\mathbb{R}} \sigma'^{-1}w(\delta)p_{\delta,\sigma'}(x^n)d\sigma' d\delta}{\int_{\sigma'>0,\delta\in\mathbb{R}} \sigma'^{-1}w(\delta)p_{0,\sigma'}(x^n)d\sigma' d\delta} \\
&= M(x^n),
\end{aligned}
$$

where we substituted $\sigma' = c\sigma$.

(ii) **[1 point]** Define $Z^n = (X_1/|X_1|, X_2/|X_1|, \ldots, X_n/|X_1|)$. Use (2) to show that, for arbitrary $X_1 \neq 0, X_2, \ldots, X_n$,

$$M(X_1, \ldots, X_n) = M(Z_1, \ldots, Z_n).$$

Solution:

For any realisation $(x_1, \ldots, x_n)$ of X^n, it follows from (i) that

$$M(x_1, \ldots, x_n) = M(x_1/|x_1|, x_2/|x_1|, \ldots, x_n/|x_1|).$$

From this, it immediately follows that $M(X_1, \ldots, X_n) = M(Z_1, \ldots, Z_n)$.

4. Fix $\sigma > 0$. Let $X_1, X_2, \ldots, X_n \sim$ i.i.d. $P_{\delta,\sigma}$. Let $X'_i = X_i/\sigma$. (i) Show that, for all $\delta \in \mathbb{R}$, the distribution of $X'_1, \ldots, X'_n$ is now i.i.d. $N(\delta, 1)$. (ii) Use (i) to show that, for each fixed δ, the distribution of Z^n is the same under $P_{\delta,\sigma}$, for all $\sigma > 0$ [for question 5. see back side!].

Solution:

(i) Consider the cumulative distribution of X'_i for arbitrary $a \in \mathbb{R}$:

$$\begin{aligned} F'_i(a) = \mathbb{P}[X'_i \leq a] &= \mathbb{P}[X_i/\sigma \leq a] \\ &= \mathbb{P}[X_i \leq \sigma a]. \end{aligned}$$

Now, let us denote p'_i for the density of X'_i:

$$\begin{aligned} p'_i(x) &= \frac{dF'_i(x)}{dx} \\ &= \frac{d}{dx}\mathbb{P}[X_i \leq \sigma x] \\ &= \frac{d}{d(\sigma x)}\mathbb{P}[X_i \leq \sigma x]\frac{d}{dx}\sigma x \\ &= p_{\delta,\sigma}(\sigma x)\sigma \\ &= \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{\sigma x}{\sigma}-\delta\right)^2}\sigma \\ &= \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\delta)^2}, \end{aligned}$$

which is indeed the density of $N(\delta, 1)$.

(ii) By (i), the distribution of $(X')^n$ is independent of σ . Then the distribution of $(Z')^n = (X'_1/|X'_1|, X'_2/|X'_1|, \ldots, X'_n/|X'_n|)$ is also independent of σ. Note that since $\sigma > 0$,

$$Z'_i = \frac{X'_i}{|X'_i|} = \frac{X_i/\sigma}{|X_i|/\sigma} = \frac{X_i}{|X_i|} = Z_i.$$

So we conclude that the distribution of Z^n is independent of σ.

As a consequence of (4)(ii), we can refer to the distribution P'_δ on Z^n without specifying the variance (the distribution does not depend on the variance of the X^n). Let p'_δ be the density of P'_δ. We can now write

$$M(X_1, \ldots, X_n) = \frac{\int_\delta w(\delta) p'_\delta(Z_1, \ldots, Z_n) d\delta}{p'_0(Z_1, \ldots, Z_n)}$$

where Z^n corresponds to X^n as above.

5. **[1 point]** Explain the following statement: even though the Bayesian universal measures in (1) are based on improper priors, and therefore do not really define probability distributions, $-\log \bar{p}_0(X^n) - [-\log \bar{p}_1(X^n)]$ *can* be interpreted as a real codelength difference between two codes.

   Solution:

   As stated,

   $$M(X_1, \ldots, X_n) = -\log \overline{p}_0(X^n) - [-\log \overline{p}_1(X^n)]$$

   is equal to

   $$M(Z_1, \ldots, Z_n) = -\log p'_0(Z_1, \ldots, Z_n) - \left[-\log p'_1(Z_1, \ldots, Z_n)\right],$$

   where $p'_1(Z_1, \ldots, Z_n) = \int_\delta w(\delta) p'_\delta(Z_1, \ldots, Z_n) d\delta$. Here p'_0 and p'_δ are probability distributions and $w(\delta)$ is a proper prior, so $M(Z_1, \ldots Z_n)$ is the actual codelength difference between p'_1 and p'_0. Note that these are codelengths for coding $Z_1, \ldots, Z_n$ (independent of σ), while the original problem was stated for $X_1, \ldots, X_n$.