Variational Bayes

Xinwei Qu

2025 年 10 月 16 日

目录

1 Variational Bayes

For a telecommunication system, we have

$$y = Hx + z, \tag{1}$$

where y is the received signal, x is the transmitted signal and $z \in \mathcal{CN}(0, \sigma^2)$. Please note that,

$$x = x_p + x_d, \tag{2}$$

where x_p is the pilot and x_d is data.

1.1 Bayes Interference

In the Rx, given the prior $p(y)$, we compute posterior distribution $p(y|x)$,

$$p(x|y) = \frac{p(x,y)}{p(y)} = \frac{\overbrace{p(y|x)}^{likelihood}\ \overbrace{p(x)}^{prior}}{\underbrace{p(y)}_{evidence}} = \frac{p(y|x)p(x)}{\int_y p(x,y)dy} \tag{3}$$

Usually, we assume the evidence is 100%, i.e., $p(y) = 1$. Hence,

$$p(x|y) \propto \overbrace{p(y|x)}^{likelihood}\ \overbrace{p(x)}^{prior} \tag{4}$$

Here, we need to choose the likelihood and the prior.

1.2 Variational Interference

However, the posterior may have no closed form, i.e., computing $p(x|y)$ is not feasible. Instead, we use a distribution Q over the symbols x to approximate $p(x|y)$, i.e.,

$$\begin{aligned} q^*(x) &= \arg\min_{q(x)\in Q} KL(q(x)||p(x|y)) \\ &= \arg\min_{q(x)\in Q} \int_x q(x)\ln\frac{q(x)}{p(x|y)}dx \\ &= \arg\min_{q(x)\in Q} -\int_x q(x)\ln\frac{p(x|y)}{q(x)}dx \end{aligned} \tag{5}$$

Here, $q^*(x)$ is the optimal $q(x)$ but $p(x|y)$ is unknown. Herefore,

$$\begin{aligned} KL(q(x)||p(x|y)) &= -\int_x q(x)\ln\frac{p(x|y)}{q(x)}dx && (6)\\ &= \int_x q(x)\ln q(x)dx - \int_x q(x)\ln p(x|y)dx && (7)\\ &= \int_x q(x)\ln q(x)dx - \int_x q(x)\ln\frac{p(x,y)}{p(y)}dx && (8)\\ &= \int_x q(x)\ln q(x)dx - \int_x q(x)\ln p(x,y)dx + \int_x q(x)\ln p(y)dx && (9)\\ &= \int_x q(x)\ln q(x)dx - \int_x q(x)\ln p(x,y)dx + \ln p(y)\int_x q(x)dx && (10)\\ &= \underbrace{\mathbb{E}_q[\ln q(x)] - \mathbb{E}_q[\ln p(x,y)]}_{-ELBO} + \ln p(y) && (11)\\ &= -ELBO(q) + \ln p(y) && (12) \end{aligned}$$

1.2.1 ELBO

Here, ELBO is Evidence Lower Bound, i.e.,

$$ELBO(q) = \mathbb{E}_q[\ln p(x,y)] - \mathbb{E}_q[\ln q(x)] \tag{13}$$

$$= \int_x q(x) \ln \frac{\overbrace{p(x,y)}^{known}}{q(x)} dx \tag{14}$$

$$= \int_x q(x) \ln \frac{p(y|x)p(x)}{q(x)} dx \tag{15}$$

(12) can be rewritten as,

$$\begin{aligned} \underbrace{\ln p(y)}_{\text{CONST}} &= ELBO(q) + \underbrace{KL(q(x)||p(x|y))}_{\geq 0} \\ &\geq ELBO(q) \end{aligned} \tag{16}$$

The minimizing KL can be taken as the maximizing ELBO, i.e.,

$$\begin{aligned} q^*(x) &= \underset{q(x)\in Q}{\arg\min}\, KL(q(x)||p(x|y)) \\ &= \underset{q(x)\in Q}{\arg\max}\, ELBO(q) \end{aligned} \tag{17}$$

1.3 Mean Field

Now, we know the problem has been simplified as the maximizing ELBO. Here, we use the mean-field assumption, i.e.,

$$\begin{aligned} q(x) &= \prod_{i=1}^{m} q_i(x_i) \\ \ln q(x) &= \sum_{i=1}^{m} \ln q_i(x_i) \\ \mathbb{E}_q[\ln q(x)] &= \sum_{i=1}^{m} \mathbb{E}_{q_i}[\ln q_i(x_i)] \end{aligned} \tag{18}$$

1.3.1 Coordinate Ascent Optimization

In $q = [q_1, q_2, \cdots, q_j, \cdots, q_m]$, we fix others to update q_j, i.e.,

$$\begin{aligned} q_j^*(x_j) &= \underset{q_j}{\arg\min}\, ELBO(q_j) \\ &= \frac{\exp\{\mathbb{E}_{q_{-j}}[\ln p(x,y)]\}}{\int_{x_j} \exp\{\mathbb{E}_{q_{-j}}[\ln p(x,y)]\} dx_j} \end{aligned} \tag{19}$$

To prove (19), we need to load (18) into (13),

$$ELBO(q) = \mathbb{E}_q[\ln p(x,y)] - \mathbb{E}_q[\ln q(x)] \tag{20}$$

$$= \int_x q(x) \ln p(x,y) dx - \left[\mathbb{E}_{q_j}[\ln q_j(x_j)] + \sum_{i\neq j} \mathbb{E}_{q_i}[\ln q_i(x_i)] \right] \tag{21}$$

Here, $\sum_{i\neq j} \mathbb{E}_{q_i}[\ln q_i(x_i)]$ can be seen as a constant because it is not related to q_j. Therefore, (21) can be simplified as ($*_{-j}$ represents the other elements except j),

$$ELBO(q) = \int_x q(x) \ln p(x,y) dx - \mathbb{E}_{q_j}[\ln q_j(x_j)] + \text{const} \tag{22}$$

$$= \int_x q(x)\ln p(x,y)dx - \int_{x_j} q(x_j)\ln q_j(x_j)dx_j + \text{const} \tag{23}$$

$$= \int_{x_j}\int_{x_{-j}} q(x_j)q(x_{-j})\ln p(x,y)dx_j dx_{-j} - \int_{x_j} q(x_j)\ln q_j(x_j)dx_j + \text{const} \tag{24}$$

$$= \int_{x_j} q(x_j)\left[\int_{x_{-j}} q(x_{-j})\ln p(x,y)dx_{-j}\right]dx_j - \int_{x_j} q(x_j)\ln q_j(x_j)dx_j + \text{const} \tag{25}$$

$$= \int_{x_j} q(x_j)\mathbb{E}_{q_{-j}}[\ln p(x,y)]dx_j - \int_{x_j} q(x_j)\ln q_j(x_j)dx_j + \text{const} \tag{26}$$

Here, we define a new distribution

$$\begin{aligned} \ln\tilde{p}_j(x_j,y) &= \mathbb{E}_{q_{-j}}[\ln p(x,y)] + const \\ \tilde{p}_j(x_j,y) &\propto \exp\{\mathbb{E}_{q_{-j}}[\ln p(x,y)]\} \end{aligned} \tag{27}$$

Here, we load (27) into (26),

$$ELBO(q) = \int_{x_j} q(x_j)\ln\tilde{p}_j(x_j,y)dx_j - \int_{x_j} q(x_j)\ln q_j(x_j)dx_j + \text{const} \tag{28}$$

$$= \int_{x_j} q(x_j)\ln\frac{\tilde{p}_j(x_j,y)}{\ln q_j(x_j)}dx_j + \text{const} \tag{29}$$

$$= -KL(q_j(x_j)||\tilde{p}_j(x_j,y)) \tag{30}$$

The KL divergence reaches the minimum when

$$\begin{aligned} q^*_{x_j} &= \tilde{p}_j(x_j,y) \\ &\propto \exp\{\mathbb{E}_{q_{-j}}[\ln p(x,y)]\} \\ &= \frac{\exp\{\mathbb{E}_{q_{-j}}[\ln p(x,y)]\}}{\int_{x_j}\exp\{\mathbb{E}_{q_{-j}}[\ln p(x,y)]\}dx_j} \end{aligned} \tag{31}$$

$\int_{x_j}\exp\{\mathbb{E}_{q_{-j}}[\ln p(x,y)]\}dx_j$是为了让总体概率为1

1.4 Algorithm Structure

The structure is given as below:

```
1: initialize q_j(x_j) for j ∈ 1, ··· , m
2: while ELBO not converge do
3:     for j ∈ 1, ··· , m do
4:         q*_{x_j} = exp{E_{q_{-j}}[ln p(x,y)]} / ∫_{x_j} exp{E_{q_{-j}}[ln p(x,y)]} dx_j
5:     end for
6:     ELBO(q)= E_q[ln p(x,y)] − E_q[ln q(x)]
7: end while
8: return q(x)
```

2 Variational Bayes in JCESD

2.1 Symbol Detection

In this section, we assume the channel is known. The target is to find the channel and the symbol to maximize the posterior, i.e.,

$$p(x;y,H,\sigma^2) = \arg\max_{x\in\Omega} p(y|x;H,\sigma^2)p(x) \tag{32}$$

Here, we use a

2.2 Joint Method

We shall look at some examples to solve this problem

$$\begin{aligned}
||y_p - \phi_p h||^2 &= (y_p - \phi_p h)^H (y_p - \phi_p h) && (33)\\
&= y_p^H y_p - y_p^H \phi_p h - (\phi_p h)^H y_p + (\phi_p h)^H (\phi_p h) && (34)\\
&= y_p^H y_p - y_p^H \phi_p h - (\phi_p h)^H y_p + h^H \phi_p^H \phi_p h && (35)
\end{aligned}$$

Here,

$$(y_p^H \phi_p h)^H = (\phi_p h)^H y_p$$

Therefore,

$$y_p^H \phi_p h + (\phi_p h)^H y_p = 2Re\{y_p^H \phi_p h\}$$

$$||y_p - \phi_p h||^2 = y_p^H y_p - 2Re\{y_p^H \phi_p h\} + h^H \phi_p^H \phi_p h$$

Now, we do the expectation for $||y_p - \phi_p h||^2$ on h,

$$\begin{aligned}
< ||y_p - \phi_p h||^2 >_h &= y_p^H y_p - 2Re\{y_p^H \phi_p u_h\} + < h^H \phi_p^H \phi_p h >_h && (36)\\
&= y_p^H y_p - 2Re\{y_p^H \phi_p u_h\} + < tr(h^H \phi_p^H \phi_p h) >_h && (37)\\
&= y_p^H y_p - 2Re\{y_p^H \phi_p u_h\} + < tr(\phi_p^H \phi_p h h^H) >_h && (38)\\
&= y_p^H y_p - 2Re\{y_p^H \phi_p u_h\} + tr(\phi_p^H \phi_p < h h^H >_h) && (39)\\
& && (40)
\end{aligned}$$

The covariance of h is,

$$\Sigma_h = < h h^H > - u_h u_h^H$$

$$< h h^H > = \Sigma_h + u_h u_h^H$$

Therefore,

$$\begin{aligned}
< ||y_p - \phi_p h||^2 >_h &= y_p^H y_p - 2Re\{y_p^H \phi_p u_h\} + tr(\phi_p^H \phi_p (\Sigma_h + u_h u_h^H)) && (41)\\
&= y_p^H y_p - 2Re\{y_p^H \phi_p u_h\} + tr(\phi_p^H \phi_p \Sigma_h) + tr(\phi_p^H \phi_p u_h u_h^H) && (42)\\
&= y_p^H y_p - 2Re\{y_p^H \phi_p u_h\} + tr(\phi_p^H \phi_p \Sigma_h) + tr(u_h^H \phi_p^H \phi_p u_h) && (43)\\
&= y_p^H y_p - 2Re\{y_p^H \phi_p u_h\} + u_h^H \phi_p^H \phi_p u_h + tr(\phi_p^H \phi_p \Sigma_h) && (44)\\
&= ||y_p - \phi_p u_h||^2 + tr(\phi_p \Sigma_h \phi_p^H) && (45)
\end{aligned}$$