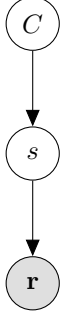


# Variational Inference for Categorization Notes

We consider the problem of categorizing stimuli with overlapping distributions

## 0.1 Generative model



$C$  is the category distribution ( $\in \{0, 1\}$ )

$C$  is the category,  $s$  is the stimulus,  $\mathbf{r}$  is the neural response and  $\tau_C$  is the precision of the  $C$  distribution

$P(C) = .5$

$s$  is the presented stimulus, a draw from the selected category distribution

$P(s|C) = \mathcal{N}(s; 0, \sigma_C^2) = \mathcal{N}(s; 0, \tau_C^{-1})$

$\mathbf{r}$  is the vector of neural responses to  $s$ :  $P(r_i|s) = \text{Poisson}(r_i; f_i(s))$

$f_i(s)$  is the tuning curve of the  $i^{\text{th}}$  neuron in response to a stimulus  $s$

Tuning curve assumptions:

- Tuning curves cover the space so  $\sum_i f_i(s)$  is independent of  $s$
- The tuning curves are Gaussian:  $f_i(s) \sim \mathcal{N}(s_i^{\text{pref}}, \sigma_{\text{tc}}^2)$

## 0.2 Inference

$$\begin{aligned}
 P(C, s|\mathbf{r}) &= \left( \prod_i P(r_i|s) \right) P(s|C) P(C) \\
 &\propto \left( \prod_i \text{Poisson}(r_i; f_i(s)) \right) \mathcal{N}(s; 0, \tau_C^{-1}) \\
 P(C, s|\mathbf{r}) &= \left( \prod_i \frac{f_i(s)^{r_i}}{r_i!} \right) \sqrt{\frac{\tau_C}{2\pi}} e^{\frac{-\tau_C s^2}{2}}
 \end{aligned} \tag{1}$$

Let

$$\begin{aligned}
\tau_C &= \begin{cases} \tau_0, & \text{if } C = 0 \\ \tau_1, & \text{if } C = 1 \end{cases} \\
&= \tau_0(1 - C) + \tau_1 C \\
&= \tau_0 - (\tau_0 - \tau_1)C \\
&= \tau_0 - C\Delta\tau
\end{aligned} \tag{2}$$

For variational inference,

$$\log P(C, s|\mathbf{r}) = \sum_i \left( -\log r_i! + r_i \log(f_i(s)) \right) + \frac{-(\tau_0 - C\Delta\tau)s^2}{2} + \frac{1}{2} \log \left( \frac{\tau_0 - C\Delta\tau}{2\pi} \right) \tag{3}$$

### 0.3 Variational approximation

Then we approximate this using a factorized distribution:

$$Q(C, s|\mathbf{r}) = Q(C|\mathbf{r})Q(s|\mathbf{r}) \tag{4}$$

$$\begin{aligned}
\log Q(C|\mathbf{r}) &= \langle \log P(C, s|\mathbf{r}) \rangle_{Q(s|\mathbf{r})} \\
&= \frac{C\Delta\tau}{2} \langle s^2 \rangle_{Q(s|\mathbf{r})} + \frac{1}{2} \log \left( \frac{\tau_0 - C\Delta\tau}{2\pi} \right) \\
&\propto \frac{C\Delta\tau}{2} \langle s^2 \rangle_{Q(s|\mathbf{r})} + \frac{1}{2} \log(\tau_0 - C\Delta\tau) \\
\log Q(C=1|\mathbf{r}) &\propto \frac{\Delta\tau}{2} \langle s^2 \rangle_{Q(s|\mathbf{r})} + \frac{1}{2} \log(\tau_0 - \Delta\tau) \\
\log Q(C=0|\mathbf{r}) &\propto \frac{1}{2} \log(\tau_0) \\
\log \frac{Q(C=1|\mathbf{r})}{Q(C=0|\mathbf{r})} &= \frac{\Delta\tau}{2} \langle s^2 \rangle_{Q(s|\mathbf{r})} + \frac{1}{2} \log \left( \frac{\tau_1}{\tau_0} \right) \\
Q(C=1|\mathbf{r}) &= \frac{1}{1 + e^{-\log \frac{Q(C=1|\mathbf{r})}{Q(C=0|\mathbf{r})}}} \\
Q(C|\mathbf{r}) &\sim \text{Bernoulli}(p) \\
p &= \frac{1}{1 + \sqrt{\frac{\tau_0}{\tau_1}} e^{-\frac{\Delta\tau}{2} \langle s^2 \rangle_{Q(s|\mathbf{r})}}} \\
Q(C|\mathbf{r}) &\sim \text{Bernoulli}(p) \\
p &= \frac{1}{1 + \sqrt{\frac{\tau_0}{\tau_1}} e^{-\frac{\Delta\tau}{2} (\frac{\eta}{\tau})^2 + \frac{1}{\tau}}}
\end{aligned} \tag{5}$$

$$\begin{aligned}
\log Q(s|\mathbf{r}) &= \langle \log P(C, s|\mathbf{r}) \rangle_{Q(C|\mathbf{r})} \\
&= \sum_i r_i \log(f_i(s)) + \frac{-(\tau_0 - \langle C \rangle_{Q(C|\mathbf{r})} \Delta\tau)}{2} s^2 \\
&\propto \frac{1}{2\sigma_{\text{tc}}^2} \sum_i r_i (s - s_i^{\text{pref}})^2 + \frac{-(\tau_0 - \langle C \rangle_{Q(C|\mathbf{r})} \Delta\tau)}{2} s^2 \\
&\sim \mathcal{N}(\mu, \tau) \\
\tau &= \sum_i \frac{r_i}{\sigma_{\text{tc}}^2} + (\tau_0 - \langle C \rangle_{Q(C|\mathbf{r})} \Delta\tau) \\
&= \sum_i \frac{r_i}{\sigma_{\text{tc}}^2} + (\tau_0 - p\Delta\tau) \\
\mu &= \frac{\sum_i \frac{r_i}{\sigma_{\text{tc}}^2} s_i^{\text{pref}}}{\sum_i \frac{r_i}{\sigma_{\text{tc}}^2} + (\tau_0 - \langle C \rangle_{Q(C|\mathbf{r})} \Delta\tau)} \\
&= \frac{\sum_i \frac{r_i}{\sigma_{\text{tc}}^2} s_i^{\text{pref}}}{\tau} \\
\eta &= \mu\tau \\
&= \sum_i \frac{r_i}{\sigma_{\text{tc}}^2} s_i^{\text{pref}} \\
Q(s|\mathbf{r}) &\sim \mathcal{N}(\eta, \tau) \\
\tau &= \sum_i \frac{r_i}{\sigma_{\text{tc}}^2} + (\tau_0 - p\Delta\tau) \\
\eta &= \sum_i \frac{r_i}{\sigma_{\text{tc}}^2} s_i^{\text{pref}}
\end{aligned} \tag{6}$$

( $\eta$  and  $\tau$  are natural parameters)

Also

$$\langle s^2 \rangle_{Q(s|\mathbf{r})} = \left(\frac{\eta}{\tau}\right)^2 + \frac{1}{\tau} \tag{7}$$

So we have update equations ( $\eta$  is a constant):

$$\begin{aligned}
\frac{dp}{dt} &= 1 - p \left( 1 + \sqrt{\frac{\tau_0}{\tau_1}} e^{-\frac{((\frac{\eta}{\tau})^2 + \frac{1}{\tau})\Delta\tau}{2}} \right) \\
\frac{d\tau}{dt} &= -\tau + \sum_i \frac{r_i}{\sigma_{\text{tc}}^2} + (\tau_0 - p\Delta\tau)
\end{aligned} \tag{8}$$

Analytic posterior:

From Qamar et al. (2013) we have

$$\begin{aligned}
\sigma_0^2 &= \frac{1}{\tau_0} \\
\sigma_1^2 &= \frac{1}{\tau_1} \\
\log \frac{P(\mathbf{r}|C=0)}{P(\mathbf{r}|C=1)} &= \frac{1}{2} \log \frac{1 + \sigma_1^2 \mathbf{a} \cdot \mathbf{r}}{1 + \sigma_0^2 \mathbf{a} \cdot \mathbf{r}} - \frac{(\sigma_1^2 - \sigma_0^2)(\mathbf{b} \cdot \mathbf{r})^2}{2(1 + \sigma_0^2 \mathbf{a} \cdot \mathbf{r})(1 + \sigma_1^2 \mathbf{a} \cdot \mathbf{r})} \quad (9) \\
P(\mathbf{r}|C=0) &= \frac{1}{1 + e^{-\left(\frac{1}{2} \log \frac{1 + \sigma_1^2 \mathbf{a} \cdot \mathbf{r}}{1 + \sigma_0^2 \mathbf{a} \cdot \mathbf{r}} - \frac{(\sigma_1^2 - \sigma_0^2)(\mathbf{b} \cdot \mathbf{r})^2}{2(1 + \sigma_0^2 \mathbf{a} \cdot \mathbf{r})(1 + \sigma_1^2 \mathbf{a} \cdot \mathbf{r})}\right)}}
\end{aligned}$$

$$\frac{1}{|\mathcal{D}|} \mathcal{L}(\theta = \{W, b\}, \mathcal{D}) = \frac{1}{|\mathcal{D}|} \sum_{i=0}^{|\mathcal{D}|} \log(P(Y = y^{(i)} | x^{(i)}, W, b)) \ell(\theta = \{W, b\}, \mathcal{D}) \quad (10)$$

$$P(C, s|\mathbf{r}) \approx Q(C|\mathbf{r})Q(s|\mathbf{r})$$

$$\log Q(C|\mathbf{r}) = \langle \log P(C, s|\mathbf{r}) \rangle_{Q(s|\mathbf{r})}$$

and

$$\log Q(s|\mathbf{r}) = \langle \log P(C, s|\mathbf{r}) \rangle_{Q(C|\mathbf{r})}$$