1 The model

There are N neurons. At each time point (in a total of T time points), it is assumed that at most one neuron can fire. Let

$$y_t \in \{0, 1, \ldots, N\}$$

indicate which neuron fired at time t (with $t = 1, \ldots, T$), with $y_t = 0$ if no neuron fired. The vector $\mathbf{y} = (y_1, \ldots, y_T)$ provides the *data*.

We construct a Hidden Markov Model (HMM) as follows. We assume that the nework can be in one of K states. Let

$$z_t \in \{1, \ldots, K\}$$

indicate the state at time t. We assume a Markov dynamics

$$P(z_{t+1} = k | z_t = l, \boldsymbol{\gamma}) = \gamma_{kl}$$

where $\boldsymbol{\gamma} = (\gamma_{11}, \gamma_{12}, \ldots, \gamma_{KK})$is the transition matrix, and the initial state z_1 is given by

$$P(z_1 = k | \boldsymbol{\pi}) = \pi_k$$

with $\pi = (\pi_1, \ldots, \pi_K)$ the initial state probabilities.

Let r_{ik} denote the probability to observe a firing of neuron i in state k (or no firing for $i = 0$). We have

$$P(y_t = i | z_t = k, \mathbf{r}) = r_{ik}$$

with $\mathbf{r} = (r_{01}, \ldots, r_{NK})$. The likelihood of the model is

$$P(\mathbf{y}|\mathbf{r}, \boldsymbol{\pi}, \boldsymbol{\gamma}) = \sum_{\mathbf{z}} P(\mathbf{y}|\mathbf{r}, \mathbf{z})\ P(\mathbf{z}|\boldsymbol{\pi}, \boldsymbol{\gamma})$$

with

$$P(\mathbf{y}|\mathbf{r}, \mathbf{z}) = \prod_{t=1}^{T} r_{y_t z_t}$$

$$P(\mathbf{z}|\boldsymbol{\pi}, \boldsymbol{\gamma}) = \pi_{z_1} \prod_{t=1}^{T-1} \gamma_{z_t z_{t+1}}$$

2 EM approach

One method to infer the parameters $\mathbf{r}$ and the latent variables $\mathbf{z}$ is to use the EM algorithm. For a general model where $\mathbf{y}$ are the data, $\mathbf{z}$ the latent variables, and $\boldsymbol{\lambda}$ the parameters, the EM works as follows. The model likelihood is

$$P(\mathbf{y}|\boldsymbol{\lambda}) = \sum_{\mathbf{z}} P(\mathbf{y}|\boldsymbol{\lambda}, \mathbf{z})\ P(\mathbf{z}|\boldsymbol{\lambda})$$

In an initialization step, the parameters $\boldsymbol{\lambda}$ are randomly initialized. Then one iterates an expectation and a maximization step.

In the expectation step, one considers the extended (non-marginalized) likelihood

$$P(\mathbf{y}, \mathbf{z}|\boldsymbol{\lambda}) = P(\mathbf{y}|\boldsymbol{\lambda}, \mathbf{z})\ P(\mathbf{z}|\boldsymbol{\lambda})$$

and computes the expectation of the log-likelihood

$$E_{\mathbf{z}}[\log P(\mathbf{y}, \mathbf{z}|\boldsymbol{\lambda})]$$

where $\mathbf{z}$ is distributed according to its posterior distribution,

$$\mathbf{z} \sim P(\mathbf{z}|\boldsymbol{\lambda}_s, \mathbf{y})$$

where $\boldsymbol{\lambda}_s$ are the values of $\boldsymbol{\lambda}$ in iteration s. This defines the expected value

$$Q(\boldsymbol{\lambda}|\boldsymbol{\lambda}_s) \equiv E_{\mathbf{z}\sim\mathbf{P}(\mathbf{z}|\boldsymbol{\lambda_s},\mathbf{y})}[\log P(\mathbf{y}, \mathbf{z}|\boldsymbol{\lambda})]$$

In the maximization step, one uptates the values of $\boldsymbol{\lambda}$:

$$\boldsymbol{\lambda}_{s+1} = \text{argmax}_{\boldsymbol{\lambda}} Q(\boldsymbol{\lambda}|\boldsymbol{\lambda}_s)$$

For the HMM, the difficult part is the E step. We have

$$Q(\mathbf{r}, \boldsymbol{\pi}, \boldsymbol{\gamma}|\mathbf{r}_s, \boldsymbol{\pi}_s, \boldsymbol{\gamma}_s) = E_{\mathbf{z}\sim P(\mathbf{z}|\mathbf{y},\mathbf{r}_s,\boldsymbol{\pi}_s,\boldsymbol{\gamma}_s)}[\log(P(\mathbf{y}|\mathbf{r}, \mathbf{z})\ P(\mathbf{z}|\boldsymbol{\pi}, \boldsymbol{\gamma}))] =$$

$$= E_{\mathbf{z}\sim P(\mathbf{z}|\mathbf{y},\mathbf{r}_s,\boldsymbol{\pi}_s,\boldsymbol{\gamma}_s)}[\log P(\mathbf{y}|\mathbf{r}, \mathbf{z}) + \log P(\mathbf{z}|\boldsymbol{\pi}, \boldsymbol{\gamma})] =$$

$$= E_{\mathbf{z}\sim P(\mathbf{z}|\mathbf{y},\mathbf{r}_s,\boldsymbol{\pi}_s,\boldsymbol{\gamma}_s)}[\sum_{t=1}^{T} \log r_{y_t z_t} + \log \pi_{z_1} + \sum_{t=1}^{T-1} \log \gamma_{z_t z_{t+1}}] =$$

$$= \sum_{t=1}^{T} E_{z_t\sim P(z_t|\mathbf{y},\mathbf{r}_s,\boldsymbol{\pi}_s,\boldsymbol{\gamma}_s)}[\log r_{y_t z_t}] + E_{z_1\sim P(z_1|\mathbf{y},\mathbf{r}_s,\boldsymbol{\pi}_s,\boldsymbol{\gamma}_s)}[\log \pi_{z_1}] +$$

$$+ \sum_{t=1}^{T-1} E_{z_t z_{t+1}\sim P(z_t z_{t+1}|\mathbf{y},\mathbf{r}_s,\boldsymbol{\pi}_s,\boldsymbol{\gamma}_s)}[\log \gamma_{z_t z_{t+1}}]$$

Thus, in the E step we need the one- and two-point posteriors

$$P(z_t|\mathbf{y}, \mathbf{r}_s, \boldsymbol{\pi}_s, \boldsymbol{\gamma}_s)$$

$$P(z_t z_{t+1}|\mathbf{y}, \mathbf{r}_s, \boldsymbol{\pi}_s, \boldsymbol{\gamma}_s)$$

The latter can be obtained with the forward-backward algorithm. One can show that the two-point posterior

$$\xi_t(k, l) = P(z_t = k, z_{t+1} = l|\mathbf{y}, \mathbf{r}_s, \boldsymbol{\pi}_s, \boldsymbol{\gamma}_s) = \frac{\alpha_t(k)\ \gamma_{kl}\ r_{y_{t+1} l}\ \beta_{t+1}(l)}{P(\mathbf{y}|\mathbf{r}, \boldsymbol{\pi}, \boldsymbol{\gamma})}$$

where the forward coefficients are defined as

$$\alpha_t(k) \equiv P(y_1, \ldots, y_t, z_t = k | \mathbf{r}, \boldsymbol{\pi}, \boldsymbol{\gamma})$$

and can be obtained recursively from

$$\alpha_1(k) = \pi_k \; r_{y_1 k}$$

with

$$\alpha_{t+1}(k) = r_{y_t k} \sum_l \; \gamma_{kl} \; \alpha_t(l)$$

and the backward coefficients are defined as

$$\beta_t(k) = P(y_{t+1}, \ldots, y_T | z_t = k, \mathbf{r}, \boldsymbol{\pi}, \boldsymbol{\gamma})$$

and can be obtained recursively from

$$\beta_T(l) \equiv 1 \; \forall l$$

with

$$\beta_{t-1}(k) = \sum_l \left(r_{y_t k} \right) \; \beta_t(l) \; \gamma_{kl}$$

The normalization constant can be shown to be

$$P(\mathbf{y} | \mathbf{r}, \boldsymbol{\pi}, \boldsymbol{\gamma}) = \sum_{kl} \xi_t(k, l)$$

The one-point posterior are simply

$$\zeta_t(k) \equiv P(z_t = k | \mathbf{y}, \mathbf{r}_s, \boldsymbol{\pi}_s, \boldsymbol{\gamma}_s) = \sum_l \xi_{kl} = \frac{\alpha_t(k) \beta_t(k)}{P(\mathbf{y} | \mathbf{r}, \boldsymbol{\pi}, \boldsymbol{\gamma})}$$

Therefore,

$$Q(\mathbf{r}, \boldsymbol{\pi}, \boldsymbol{\gamma} | \mathbf{r}_s, \boldsymbol{\pi}_s, \boldsymbol{\gamma}_s) = \sum_{t=1}^{T} (\sum_k \zeta_t(k) \log r_{y_t k}) + \sum_k \zeta_1(k) \log \pi_k + \sum_{t=1}^{T-1} (\sum_{kl} \xi_t(k, l) \log \gamma_{kl}$$

By maximizing Q over $\boldsymbol{\pi}, \boldsymbol{\gamma}, \mathbf{r}$ one obtains the updates $\mathbf{r}_{s+1}, \boldsymbol{\pi}_{s+1}, \boldsymbol{\gamma}_{s+1}$. The maximization can be done analytically very simply by taking derivatives (and accounting for constraints; for instance, since $\sum_i r_{ik} = 1$, one has to add the corresponding Lagrange multiplier, etc.). One obtains the optimal values

$$\gamma_{kl} = \frac{\sum_{t=1}^{T-1} \xi_{kl}(t)}{\sum_{t=1}^{T-1} \zeta_k(t)}$$

$$\pi_k = \zeta_1(k)$$

$$r_{ik} = \frac{\sum_{t=1}^{T} \zeta_t(k) \delta(y_t = i)}{\sum_{i=0}^{N} \sum_{t=1}^{T} \zeta_t(k) \delta(y_t = i)}$$

Gibbs approach

In the Gibbs approach, one can sample the full posterior

$$P(\mathbf{r}, \boldsymbol{\pi}, \boldsymbol{\gamma}, \mathbf{z}|\mathbf{y}) = P(\mathbf{y}|\mathbf{r}, \mathbf{z})P(\mathbf{z}|\boldsymbol{\pi}, \boldsymbol{\gamma})P(\boldsymbol{\pi})P(\boldsymbol{\gamma})P(\mathbf{r})\frac{1}{P(\mathbf{y})}$$

The priors for the parameters can be taken conjugate. The prior for the initial state probabilities is a Dirichlet with parameters $\mathbf{v} = (v_1, \ldots, v_K)$ (for $v_k = 1 \forall k$ one obtains the uniform prior).

$$\pi \sim dir(\mathbf{v})$$

The prior for each row of the transition matrix, $\boldsymbol{\gamma}_k = (\gamma_{k1}, \ldots, \gamma_{kl})$, is a Dirichlet with parameters $\mathbf{u}_k = (u_{k1}, \ldots, u_{kK})$ (for $u_{kl} = 1 \forall l$ one obtains the uniform prior).

$$\boldsymbol{\gamma}_k \sim dir(\mathbf{u}_k)$$

The prior for each column of $\mathbf{r}$, $\boldsymbol{r}_k = (r_{0k}, \ldots, \gamma_{Nk})$, is a Dirichlet with parameters $\mathbf{w}_k = (w_{0k}, \ldots, w_{Nk})$ (for $w_{ik} = 1 \forall i$ one obtains the uniform prior).

$$\boldsymbol{r}_k \sim dir(\mathbf{w}_k)$$

The conditional posteriors of $\mathbf{r}, \boldsymbol{\pi}, \boldsymbol{\gamma}$ will still be Dirichlets with updates parameters. In fact, the Gibbs samplig alternates the following steps

1. Update
$$v'_k = v_k + \delta(z_1 = k)$$
and sample
$$\pi \sim dir(\mathbf{v}')$$

2. Update
$$u'_{kl} \to u_{kl} + \sum_{t=1}^{T-1} \delta(z_t = k)\delta(z_{t+1} = l)$$
and sample
$$\boldsymbol{\gamma}_k \sim dir(\mathbf{u}'_k)$$

3. Update
$$w_{ik} \to w_{ik} + \sum_t \delta(z_t = k)\delta(y_t = i)$$
and sample
$$\boldsymbol{r}_k \sim dir(\mathbf{w}'_k)$$

4. For each t, sample z_t from a categorical distribution,
$$P(z_t = k|z_{t+1} = l, z_{t-1} = m) = \eta_{k,l,m}$$
with
$$\eta_k = \frac{r_{y_t k}\gamma_{mk}\gamma_{kl}}{\sum_m r_{y_t k}\gamma_{mk}\gamma_{kl}}$$