Molecular Classification

BIOF2014

Problem

We want to classify tumours based on their molecular characteristics, as determined by RNA expression profiles. Each expression profile consists of the expression levels of selected genes, measured as counts of molecules.

Some cancer types can be classified reliably into molecular classes, and these molecule classes may reflect different cellular origins of the tumours. For example, the pediatric brain tumour known as medulloblastoma can be molecularly classified into different subtypes.

Model

Given training data with expression profiles $\mathbf{X} = [\boldsymbol{x}_i^\top]$ and class labels $\boldsymbol{y} = [y_i]$ for $i \in \{1 \ldots N\}$, we want to predict the unknown label $\tilde{y}$ of a new sample with expression profile $\tilde{\boldsymbol{x}}$.

Each expression profile consists the detected counts of transcript molecules of J genes, so $\boldsymbol{x}_i, \tilde{\boldsymbol{x}} \in \mathcal{N}_0^J$. Suppose there are K classes, so $y_i, \tilde{y} \in \{1 \ldots K\}$.

Define

$$m_i = \sum_j x_{ij}, \quad n_k = \sum_i I(y_i = k), \quad s_{kj} = \sum_i x_{ij} I(y_i = k).$$

Define our model as follows:

$$\begin{aligned}
\boldsymbol{X}_i = \boldsymbol{x} \mid Y_i = y,\, \boldsymbol{\eta}_y \;&\sim\; \text{Multinomial}\left(m_i \mid \boldsymbol{\eta}_y\right) \\
\boldsymbol{\eta}_k &\sim \text{Dirichlet}\left(\boldsymbol{c}_k\right) \\
Y_i &\sim \boldsymbol{\theta} \\
\boldsymbol{\theta} &\sim \text{Dirichlet}\left(\boldsymbol{d}\right),
\end{aligned}$$

where $\boldsymbol{c}_k \in \mathcal{R}_{\geq 0}^J$ and $\boldsymbol{d} \in \mathcal{R}_{\geq 0}^K$ are hyperparameters.

Model fitting

Outline

Model fitting involves using the observed data $(\mathbf{X}, \boldsymbol{y})$ to learn the model parameters $((\boldsymbol{\eta}_k), \boldsymbol{\theta})$. The hyperparameters $((\boldsymbol{c}_k), \boldsymbol{d})$ can be set based on prior knowledge or data.

Therefore, the objective of model fitting (i.e. training) is to estimate the posterior distributions of the parameters.

For $\boldsymbol{\theta}$, we can apply Bayes' theorem and get

$$p(\boldsymbol{\theta} \mid \boldsymbol{y}) = \frac{p(\boldsymbol{y} \mid \boldsymbol{\theta})\ p(\boldsymbol{\theta})}{p(\boldsymbol{y})}.$$

For $k \in \{1, \dots, K\}$, we need to solve

$$p(\boldsymbol{\eta}_k \mid \mathbf{X}, \boldsymbol{y}) = \frac{p(\mathbf{X} \mid \boldsymbol{\eta}_k, \boldsymbol{y})\ p(\boldsymbol{\eta}_k \mid \boldsymbol{y})}{p(\mathbf{X} \mid \boldsymbol{y})},$$

which can be seen as the application of a special case of the Bayes' theorem.

Derivation

Let's solve $p(\boldsymbol{\theta} \mid \boldsymbol{y})$.

$$\begin{aligned}
p(\boldsymbol{y} \mid \boldsymbol{\theta})\ p(\boldsymbol{\theta}) &= \left(\prod_i^N p(y_i \mid \boldsymbol{\theta})\right)\ p(\boldsymbol{\theta}) \\
&= \left(\prod_i^N \theta_{y_i}\right)\ \left(B(\boldsymbol{d})^{-1} \prod_k \theta_k^{d_k - 1}\right) \\
&= \left(\prod_i^N \prod_k^K (\theta_k)^{I(y_i = k)}\right)\ \left(B(\boldsymbol{d})^{-1} \prod_k \theta_k^{d_k - 1}\right) \\
&= \left(\prod_k^K (\theta_k)^{\sum_i^N I(y_i = k)}\right)\ \left(B(\boldsymbol{d})^{-1} \prod_k \theta_k^{d_k - 1}\right) \\
&= B(\boldsymbol{d})^{-1} \left(\prod_k^K (\theta_k)^{d_k + n_k - 1}\right).
\end{aligned}$$

Now, let's solve for

$$p(\boldsymbol{y}) = \int_\Theta p(\boldsymbol{y} \mid \boldsymbol{\theta})\ p(\boldsymbol{\theta})\ d\boldsymbol{\theta}$$

From the definition of Dirichlet($\boldsymbol{x} \mid \boldsymbol{\alpha}$), we know that

$$\int_{\mathcal{X}} \left(B(\boldsymbol{\alpha})^{-1} \prod_k x_k^{\alpha_k - 1} \right) d\boldsymbol{x} = 1$$
$$B(\boldsymbol{\alpha}) = \int_{\mathcal{X}} \left(\prod_k \theta_k^{\alpha_k - 1} \right) d\boldsymbol{x},$$

for any $\boldsymbol{\alpha}$ such that $\alpha_k > 0$.

Define $\boldsymbol{d'} = [d'_k]$ where $d'_k = d_k + n_k$. Then,

$$\int_{\Theta} \prod_k^K (\theta_k)^{d_k + n_k - 1} \, d\boldsymbol{\theta} = \int_{\Theta} \prod_k^K (\theta_k)^{d'_k - 1} \, d\boldsymbol{\theta} = B(\boldsymbol{d'}).$$

Therefore,

$$\begin{aligned} p(\boldsymbol{y}) = \int_{\Theta} p(\boldsymbol{y} \mid \boldsymbol{\theta}) \, p(\boldsymbol{\theta}) \, d\boldsymbol{\theta} \quad &= \int_{\Theta} B(\boldsymbol{d})^{-1} \left(\prod_k^K (\theta_k)^{d_k + n_k - 1} \right) d\boldsymbol{\theta} = B(\boldsymbol{d})^{-1} \int_{\Theta} \left(\prod_k^K (\theta_k)^{d_k + n_k - 1} \right) d\boldsymbol{\theta} \\ &= B(\boldsymbol{d})^{-1} B(\boldsymbol{d'}). \end{aligned}$$

Now, we are ready to solve for the posterior:

$$p(\boldsymbol{\theta} \mid \boldsymbol{y}) = \frac{p(\boldsymbol{y} \mid \boldsymbol{\theta}) \, p(\boldsymbol{\theta})}{p(\boldsymbol{y})} \quad = \frac{B(\boldsymbol{d})^{-1} \left(\prod_k^K (\theta_k)^{d_k + n_k - 1} \right)}{B(\boldsymbol{d})^{-1} B(\boldsymbol{d'})} = B(\boldsymbol{d'})^{-1} \prod_k^K (\theta_k)^{d'_k - 1}.$$

Therefore, the posterior of $\boldsymbol{\theta}$ is

$$p(\boldsymbol{\theta} \mid \boldsymbol{y}) = \text{Dirichlet}(\boldsymbol{\theta} \mid \boldsymbol{d'}),$$

where $\boldsymbol{d'} = \left[d_k + \sum_i I(y_i = k) \right]$.

Next, let's solve, for any $k \in \{1, \ldots, K\}$,

$$p(\boldsymbol{\eta}_k \mid \mathbf{X}, \boldsymbol{y}) = \frac{p(\mathbf{X} \mid \boldsymbol{\eta}_k, \boldsymbol{y}) \, p(\boldsymbol{\eta}_k \mid \boldsymbol{y})}{p(\mathbf{X} \mid \boldsymbol{y})}.$$

As before, we will start with the numerator:

$$
\begin{aligned}
p\left(\mathbf{X} \mid \boldsymbol{\eta}_k, \boldsymbol{y}\right)\, p\left(\boldsymbol{\eta}_k \mid \boldsymbol{y}\right) &= \left(\prod_i^N p\left(\boldsymbol{x}_i \mid \boldsymbol{\eta}_{y_i}, y_i\right)\right)\, p\left(\boldsymbol{\eta}_k\right) \\
&= \left(\prod_i^N \Gamma(M_i+1) \prod_j^J \Gamma(x_{ij}+1)^{-1} \eta_{y_i j}^{x_{ij}}\right) \left(B(\boldsymbol{c}_k)^{-1} \prod_j^J \eta_{kj}^{c_{kj}-1}\right) \\
&= \left(\prod_i^N \Gamma(M_i+1)\right) \left(\prod_i^N \prod_j^J \Gamma(x_{ij}+1)^{-1}\right) \left(\prod_i^N \prod_j^J \eta_{y_i j}^{x_{ij}}\right) \left(B(\boldsymbol{c}_k)^{-1} \prod_j^J \eta_{kj}^{c_{kj}-1}\right) \\
&= g_1(\mathbf{X}) \left(\prod_i^N \prod_j^J \eta_{y_i j}^{x_{ij}}\right) \left(B(\boldsymbol{c}_k)^{-1} \prod_j^J \eta_{kj}^{c_{kj}-1}\right) \\
&= g_1(\mathbf{X}) \left(\prod_{i:y_i \neq k}^N \prod_j^J \eta_{y_i j}^{x_{ij}}\right) \left(\prod_{i:y_i = k}^N \prod_j^J \eta_{kj}^{x_{ij}}\right) \left(B(\boldsymbol{c}_k)^{-1} \prod_j^J \eta_{kj}^{c_{kj}-1}\right) \\
&= g_1(\mathbf{X})\, g_2(\mathbf{X}, \boldsymbol{y}, k) \left(\prod_{i:y_i = k}^N \prod_j^J \eta_{kj}^{x_{ij}}\right) \left(B(\boldsymbol{c}_k)^{-1} \prod_j^J \eta_{kj}^{c_{kj}-1}\right) \\
&= g_1(\mathbf{X})\, g_2(\mathbf{X}, \boldsymbol{y}, k) \left(\prod_j^J \eta_{kj}^{\sum_i x_{ij}\, I(y_i=k)}\right) \left(B(\boldsymbol{c}_k)^{-1} \prod_j^J \eta_{kj}^{c_{kj}-1}\right) \\
&= g_1(\mathbf{X})\, g_2(\mathbf{X}, \boldsymbol{y}, k)\, B(\boldsymbol{c}_k)^{-1} \left(\prod_j^J \eta_{kj}^{\sum_i x_{ij}\, I(y_i=k)}\right) \prod_j^J \eta_{kj}^{c_{kj}-1} \\
&= g_1(\mathbf{X})\, g_2(\mathbf{X}, \boldsymbol{y}, k)\, B(\boldsymbol{c}_k)^{-1} \prod_j^J \eta_{kj}^{c_{kj}+s_{kj}-1}
\end{aligned}
$$

where

$$
\begin{aligned}
g_1(\mathbf{X}) &= \left(\prod_i^N \Gamma(M_i+1)\right) \left(\prod_i^N \prod_j^J \Gamma(x_{ij}+1)^{-1}\right) \\
g_2(\mathbf{X}, \boldsymbol{y}, k) &= \left(\prod_{i:y_i \neq k}^N \prod_j^J \eta_{y_i j}^{x_{ij}}\right).
\end{aligned}
$$

Now, we solve for the denominator:

$$\begin{aligned}
p\left(\mathbf{X},\boldsymbol{y}\right) &= \int_{\mathrm{H}} p\left(\mathbf{X}\mid\boldsymbol{\eta}_k,\boldsymbol{y}\right)\, p\left(\boldsymbol{\eta}_k\mid\boldsymbol{y}\right)\, d\boldsymbol{\eta}_k \\
&= \int_{\mathrm{H}} g_1(\mathbf{X})\, g_2(\mathbf{X},\boldsymbol{y},k)\, B(\boldsymbol{c}_k)^{-1} \prod_j^J \eta_{yj}^{c_{yj}+s_{yj}-1}\, d\boldsymbol{\eta}_k \\
&= g_1(\mathbf{X})\, g_2(\mathbf{X},\boldsymbol{y},k)\, B(\boldsymbol{c}_k)^{-1} \int_{\mathrm{H}} \prod_j^J \eta_{yj}^{c_{yj}+s_{yj}-1}\, d\boldsymbol{\eta}_k \\
&= g_1(\mathbf{X})\, g_2(\mathbf{X},\boldsymbol{y},k)\, B(\boldsymbol{c}_k)^{-1}\, B(\boldsymbol{c}_k'),
\end{aligned}$$

where $\boldsymbol{c}_k' = [c_{kj}']$ and $c_{kj}' = c_{kj} + s_{kj}$.

Putting the numerator and denominator together, we get

$$\begin{aligned}
p\left(\boldsymbol{\eta}_k\mid\mathbf{X},\boldsymbol{y}\right) = \frac{p\left(\mathbf{X}\mid\boldsymbol{\eta}_k,\boldsymbol{y}\right)\, p\left(\boldsymbol{\eta}_k\mid\boldsymbol{y}\right)}{p\left(\mathbf{X}\mid\boldsymbol{y}\right)} &= \frac{g_1(\mathbf{X})\, g_2(\mathbf{X},\boldsymbol{y},k)\, B(\boldsymbol{c}_k)^{-1} \prod_j^J \eta_{kj}^{c_{kj}'-1}}{g_1(\mathbf{X})\, g_2(\mathbf{X},\boldsymbol{y},k)\, B(\boldsymbol{c}_k)^{-1}\, B(\boldsymbol{c}_k'),} \\
&= B(\boldsymbol{c}_k')^{-1} \prod_j^J \eta_{kj}^{c_{kj}'-1}.
\end{aligned}$$

Therefore, for each k, the posterior of $\boldsymbol{\eta}_k$ conditional on $\boldsymbol{y}$ is

$$p\left(\boldsymbol{\eta}_k\mid\mathbf{X},\boldsymbol{y}\right) = \mathrm{Dirichlet}\left(\boldsymbol{\eta}_k\mid\boldsymbol{c}'\right),$$

where $\boldsymbol{c}_k' = \left[c_{kj} + \sum_i x_{ij} I(y_i = k)\right]$.

Model prediction

Outline

To use our model for classification, we need to derive the posterior predictive distribution

$$p\left(\tilde{y}\mid\tilde{\boldsymbol{x}},\mathbf{X},\boldsymbol{y}\right) = \frac{p\left(\tilde{\boldsymbol{x}}\mid\tilde{y},\mathbf{X},\boldsymbol{y}\right)\, p\left(\tilde{y}\mid\boldsymbol{y}\right)}{p\left(\tilde{\boldsymbol{x}}\mid\mathbf{X},\boldsymbol{y}\right)}.$$

In turn, we need to derive other posterior predictive distributions

$$\begin{aligned}
p\left(\tilde{\boldsymbol{x}}\mid\tilde{y},\mathbf{X},\boldsymbol{y}\right) &= \int p\left(\tilde{\boldsymbol{x}}\mid\tilde{y},\boldsymbol{\eta}_{\tilde{y}}\right)\, p\left(\boldsymbol{\eta}_{\tilde{y}}\mid\mathbf{X},\boldsymbol{y}\right)\, d\boldsymbol{\eta}_{\tilde{y}} \\
p\left(\tilde{y}\mid\boldsymbol{y}\right) &= \int p\left(\tilde{y}\mid\boldsymbol{\theta}\right)\, p\left(\boldsymbol{\theta}\mid\boldsymbol{y}\right)\, d\boldsymbol{\theta}.
\end{aligned}$$

Given the above quantities, we can then derive

$$p\left(\tilde{x}\mid\mathbf{X},\boldsymbol{y}\right) = \sum_{k=1}^K p\left(\tilde{\boldsymbol{x}}\mid\tilde{Y}=k,\mathbf{X},\boldsymbol{y}\right)\, p\left(\tilde{Y}=k\mid\boldsymbol{y}\right).$$

Derivation

Let's begin by deriving the posterior predictive distribution for $\tilde{y}$.

$$\begin{aligned}
p\left(\tilde{y} \mid \boldsymbol{y}\right) &= \int_{\Theta} p\left(\tilde{y} \mid \boldsymbol{\theta}\right)\, p\left(\boldsymbol{\theta} \mid \boldsymbol{y}\right)\, d\boldsymbol{\theta} \\
&= \int_{\Theta} \theta_{\tilde{y}}\, p\left(\boldsymbol{\theta} \mid \boldsymbol{y}\right)\, d\boldsymbol{\theta} \\
&= \int_0^1 \int_{\Theta_{-\tilde{y}}} \theta_{\tilde{y}}\, p\left(\boldsymbol{\theta} \mid \boldsymbol{y}\right)\, d\boldsymbol{\theta}_{-\tilde{y}}\, d\theta_{\tilde{y}} \\
&= \int_0^1 \theta_{\tilde{y}} \int_{\Theta_{-\tilde{y}}} p\left(\boldsymbol{\theta} \mid \boldsymbol{y}\right)\, d\boldsymbol{\theta}_{-\tilde{y}}\, d\theta_{\tilde{y}} \\
&= \int_0^1 \theta_{\tilde{y}} \int_{\Theta_{-\tilde{y}}} p\left(\theta_{\tilde{y}}, \boldsymbol{\theta}_{-\tilde{y}} \mid \boldsymbol{y}\right)\, d\boldsymbol{\theta}_{-\tilde{y}}\, d\theta_{\tilde{y}} \\
&= \int_0^1 \theta_{\tilde{y}}\, p\left(\theta_{\tilde{y}} \mid \boldsymbol{y}\right)\, d\theta_{\tilde{y}} \\
&= \mathrm{E}_{p\left(\theta_{\tilde{y}} \mid \boldsymbol{y}\right)}\left(\theta_{\tilde{y}}\right)
\end{aligned}$$

As we showed above, $p\left(\boldsymbol{\theta} \mid \boldsymbol{y}\right)$ is $\text{Dirichlet}(\theta \mid \boldsymbol{d}')$. Therefore,

$$p\left(\tilde{y} \mid \boldsymbol{y}\right) = \mathrm{E}_{p\left(\theta_{\tilde{y}} \mid \boldsymbol{y}\right)}\left(\theta_{\tilde{y}}\right) = \frac{d'_{\tilde{y}}}{\sum_k d'_k}.$$

Since this quantity will be useful, we define $\tilde{\theta}_k \ = \ p\left(\tilde{Y} = k \mid \boldsymbol{y}\right)$ for all k.

Now, let's derive the posterior predictive distribution for $\tilde{\boldsymbol{x}}$. Define $\tilde{m} = \sum_j \tilde{x}_j$.

$$\begin{aligned}
p\left(\tilde{\boldsymbol{x}} \mid \tilde{y}, \mathbf{X}, \boldsymbol{y}\right) &= \int p\left(\tilde{\boldsymbol{x}} \mid \tilde{y}, \boldsymbol{\eta}_{\tilde{y}}\right)\, p\left(\boldsymbol{\eta}_{\tilde{y}} \mid \mathbf{X}, \boldsymbol{y}\right)\, d\boldsymbol{\eta}_{\tilde{y}} \\
&= \int_{\mathrm{H}} \left(\Gamma(\tilde{m}+1) \prod_j^J \Gamma(\tilde{x}_j+1)^{-1}\, \eta_{\tilde{y}j}^{\tilde{x}_j}\right) \left(B(\boldsymbol{c}'_{\tilde{y}})^{-1} \prod_j^J \eta_{\tilde{y}j}^{c'_{\tilde{y}j}-1}\right) d\boldsymbol{\eta}_{\tilde{y}} \\
&= \int_{\mathrm{H}} \Gamma(\tilde{m}+1) \left(\prod_j^J \Gamma(\tilde{x}_j+1)^{-1}\right) \left(\prod_j^J \eta_{\tilde{y}j}^{\tilde{x}_j}\right) B(\boldsymbol{c}'_{\tilde{y}j})^{-1} \prod_j^J \eta_{\tilde{y}j}^{c'_{\tilde{y}j}-1} d\boldsymbol{\eta}_{\tilde{y}} \\
&= \Gamma(\tilde{m}+1) \left(\prod_j^J \Gamma(\tilde{x}_j+1)^{-1}\right) B(\boldsymbol{c}'_{\tilde{y}})^{-1} \int_{\mathrm{H}} \left(\prod_j^J \eta_{\tilde{y}j}^{\tilde{x}_j}\right) \left(\prod_j^J \eta_{\tilde{y}j}^{c'_{\tilde{y}j}-1}\right) d\boldsymbol{\eta}_{\tilde{y}} \\
&= \Gamma(\tilde{m}+1) \left(\prod_j^J \Gamma(\tilde{x}_j+1)^{-1}\right) B(\boldsymbol{c}'_{\tilde{y}})^{-1} \int_{\mathrm{H}} \left(\prod_j^J \eta_{\tilde{y}j}^{c'_{\tilde{y}j}+\tilde{x}_j-1}\right) d\boldsymbol{\eta}_{\tilde{y}} \\
&= \Gamma(\tilde{m}+1) \left(\prod_j^J \Gamma(\tilde{x}_j+1)^{-1}\right) B(\boldsymbol{c}'_{\tilde{y}})^{-1}\, B(\boldsymbol{c}''_{\tilde{y}}),
\end{aligned}$$

where $c''_{\tilde{y}j} = c'_{\tilde{y}j} + \tilde{x}_j$.

Now, let's derive the denominator:

$$
\begin{aligned}
p\left(\tilde{x} \mid \mathbf{X}, \boldsymbol{y}\right) &= \sum_{k=1}^{K} p\left(\tilde{\boldsymbol{x}} \mid \tilde{Y} = k, \mathbf{X}, \boldsymbol{y}\right)\, p\left(\tilde{Y} = k \mid \boldsymbol{y}\right) \\
&= \sum_{k=1}^{K} \Gamma(\tilde{m}+1) \left(\prod_{j}^{J} \Gamma(\tilde{x}_j + 1)^{-1}\right) B(\boldsymbol{c}'_k)^{-1}\, B(\boldsymbol{c}''_k)\, \tilde{\theta}_k \\
&= \Gamma(\tilde{m}+1) \left(\prod_{j}^{J} \Gamma(\tilde{x}_j + 1)^{-1}\right) \sum_{k=1}^{K} B(\boldsymbol{c}'_k)^{-1}\, B(\boldsymbol{c}''_k)\, \tilde{\theta}_k.
\end{aligned}
$$

We are now finally ready to derive the posterior predictive distribution of $\tilde{y}$:

$$
\begin{aligned}
p\left(\tilde{y} \mid \tilde{\boldsymbol{x}}, \mathbf{X}, \boldsymbol{y}\right) &= \frac{p\left(\tilde{\boldsymbol{x}} \mid \tilde{y}, \mathbf{X}, \boldsymbol{y}\right)\, p\left(\tilde{y} \mid \boldsymbol{y}\right)}{p\left(\tilde{\boldsymbol{x}} \mid \mathbf{X}, \boldsymbol{y}\right)} \\
&= \frac{\Gamma(\tilde{m}+1) \left(\prod_{j}^{J} \Gamma(\tilde{x}_j + 1)^{-1}\right) B(\boldsymbol{c}'_{\tilde{y}})^{-1}\, B(\boldsymbol{c}''_{\tilde{y}})\, \tilde{\theta}_{\tilde{y}}}{\Gamma(\tilde{m}+1) \left(\prod_{j}^{J} \Gamma(\tilde{x}_j + 1)^{-1}\right) \sum_{k=1}^{K} B(\boldsymbol{c}'_k)^{-1}\, B(\boldsymbol{c}''_k)\, \tilde{\theta}_k} \\
&= \frac{B(\boldsymbol{c}'_{\tilde{y}})^{-1}\, B(\boldsymbol{c}''_{\tilde{y}})\, \tilde{\theta}_{\tilde{y}}}{\sum_{k=1}^{K} B(\boldsymbol{c}'_k)^{-1}\, B(\boldsymbol{c}''_k)\, \tilde{\theta}_k}.
\end{aligned}
$$

Questions

1. What is the dimension of $\boldsymbol{\theta}$? What are the constraints for each element of $\boldsymbol{\theta}$?
2. What is the dimension of each $\boldsymbol{\eta}_k$? What are the constraints for each element of $\boldsymbol{\eta}_k$?
3. How many individual observed data elements are provided to the model?
4. How many individual parameters and hyperparameters are in the model?