External Use of Rate Matrices

Mario Stanke

February 2021

1 Continuous-Time Markov Chains

The folder `vert-fly-yeast-LogReg-M3` contains the parameters of a ClaMSA model with $M = 3$ rate matrices that was trained on vertebrate, fly and yeast alignments. `rates-Q.txt` contains the $M \cdot 64^2 = 12288$ float values for rate matrices $Q^{(1)}, Q^{(2)}, Q^{(3)}$. `rates-pi.txt` contains the $64 \cdot M = 192$ float values for corresponding stationary distributions $\pi^{(1)}, \pi^{(2)}, \pi^{(3)}$ and $\pi_1^{(m)} + \cdots + \pi_{64}^{(m)} = 1$. They can also be inferred from the the rate matrices and satisfy

$$\pi^{(m)} Q^{(m)} = 0$$

Let L_i^m be the log-likelihood of the i-th codon alignment column under the m-th model ($m = 1, \ldots, M$).

2 Prediction Layer

The 'sequence layers' of `vert-fly-yeast-LogReg-M3` are a simple logistic regression model with two output classes as described in the manuscript:

```
means = mean(L, axis = 1)
y = Dense(2, activation = "softmax")(means)
```

These prediction layers have only $2(M+1) = 8$ parameters that are stored in file `Theta.txt`. Let $\Theta \in \mathbb{R}^{M \times 2}$ and $\theta \in \mathbb{R}^2$ be the matrix and bias vector stored in `Theta.txt` the dimension size 2 is the number of output classes.

Let

$$\bar{L}_m = \frac{1}{\ell} \sum_{i=1}^{\ell} L_i^m \qquad (m = 1, \ldots, M),$$

$\bar{L} := (\bar{L}_1, \ldots, \bar{L}_M)$ and

$$\mathbf{z} := \theta + \bar{L}\Theta \in \mathbb{R}^2.$$

Let

$$\mathbf{y} = (y_0, y_1) = \mathbf{g}(\mathbf{z}) = \frac{\mathbf{1}}{\sum_{\mathbf{k=1}}^{\mathbf{2}} \mathbf{e}^{\mathbf{z_k}}} \begin{pmatrix} e^{z_1} \\ e^{z_2} \end{pmatrix},$$

i.e. $\mathbf{g}$ is the softmax function.

Then $y_1 = \sigma(z_2 - z_1)$ is the predicted probability of the positive class, where $\sigma(r) = 1/(1 + \exp(-r))$ is the logistic sigmoid function.

Therefore, after the parameter transformation $w_0 := \theta_2 - \theta_1$ and $w_m := \Theta_{m,2} - \Theta_{m,1}$ for $m = 1..M$, we obtain $M + 1 = 4$ free parameters and the predicted probability of class 1

$$y_1 = \sigma(w_0 + w_1 \bar{L}_1 + \cdots + w_M \bar{L}_M).$$