M526: Final

Dominic Bair

May 15, 2023

1. First, we implement the normalized forward filtering algorithm on our corrupted data using the following model:

$$s_1 \sim \text{Categorical}_{\sigma_{1:M}}(\tilde{\rho})$$
$$s_n | s_{n-1} \sim \text{Categorical}_{\sigma_{1:M}}(\pi), \qquad\qquad n = 2, \cdots, N$$
$$w_n | s_n \sim \text{Poisson}(\phi_{s_n}), \qquad\qquad n = 1, \cdots, N$$

Where $\tilde{\rho} = \begin{bmatrix} 1/5 \\ 1/5 \\ 1/5 \\ 1/5 \\ 1/5 \end{bmatrix}$, $\pi = \begin{bmatrix} 25/26 & 1/26 & 0 & 0 & 0 \\ 1/27 & 25/27 & 1/27 & 0 & 0 \\ 0 & 1/27 & 25/27 & 1/27 & 0 \\ 0 & 0 & 1/27 & 25/27 & 1/27 \\ 0 & 0 & 0 & 1/26 & 25/26 \end{bmatrix}$, and $\tilde{\phi} = \begin{bmatrix} 8 \\ 15 \\ 24 \\ 31 \\ 42 \end{bmatrix}$.

We find the (marginal) log–likelihood of our data to be -7.7447×10^3.

2. Now we that we have computed our filters, we implement the Viterbi Algorithm on our data to classify the category each data point belongs in. See Figure 1
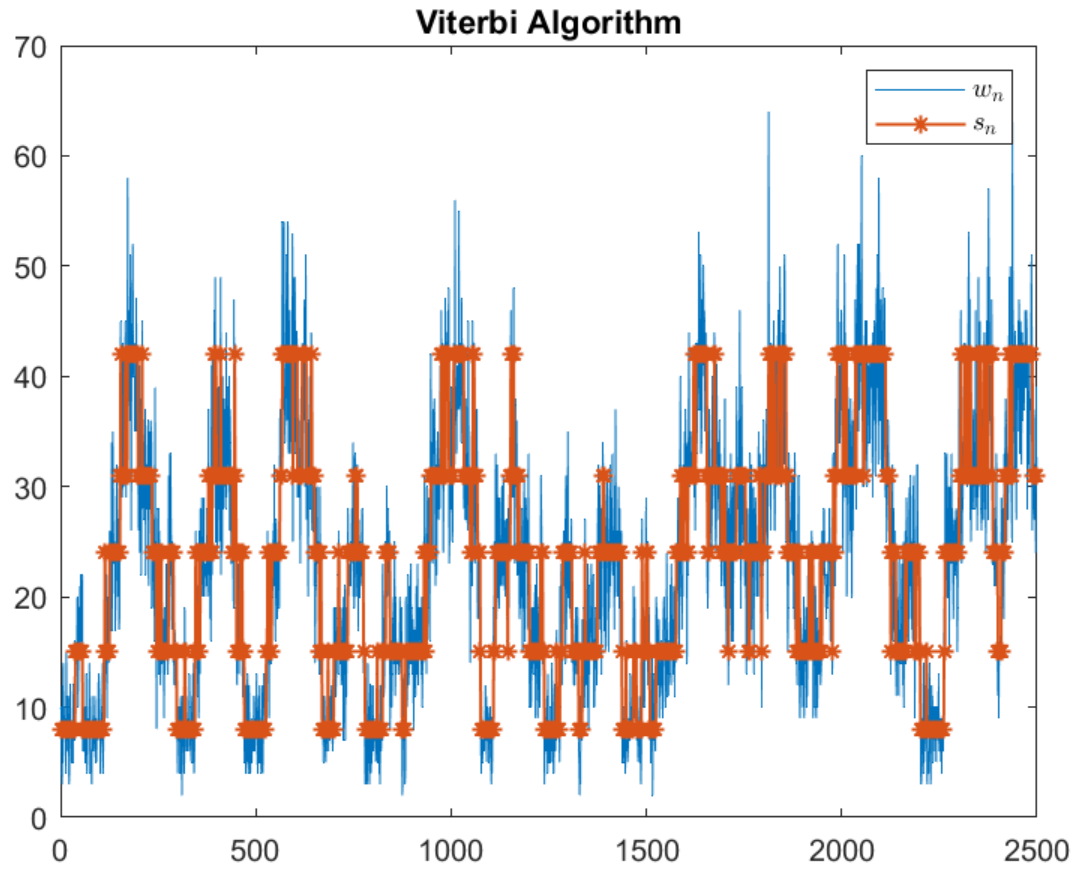
Figure 1: A visualization of the Viterbi Algorithm. We fist classify each datum into a category, then plot it at the expected number of occurrences of its prescribed category. We see that the algorithm closely matches the data.

3. We now create a Bayesian Model to estimate all model parameters.

$$\tilde{\rho} \sim \text{Dirichlet}_{\sigma_{1:M}}(\alpha\tilde{\beta})$$
$$\tilde{\pi}_{\sigma_1} \sim \text{Dirichlet}_{\sigma_{1:M}}(\alpha\tilde{\beta})$$
$$\tilde{\pi}_{\sigma_2} \sim \text{Dirichlet}_{\sigma_{1:M}}(\alpha\tilde{\beta})$$
$$\tilde{\pi}_{\sigma_3} \sim \text{Dirichlet}_{\sigma_{1:M}}(\alpha\tilde{\beta})$$
$$\tilde{\pi}_{\sigma_4} \sim \text{Dirichlet}_{\sigma_{1:M}}(\alpha\tilde{\beta})$$
$$\tilde{\pi}_{\sigma_5} \sim \text{Dirichlet}_{\sigma_{1:M}}(\alpha\tilde{\beta})$$
$$\phi_{\sigma_1} \sim \text{Gamma}(A_1, B_1)$$
$$\phi_{\sigma_2} \sim \text{Gamma}(A_2, B_2)$$
$$\phi_{\sigma_3} \sim \text{Gamma}(A_3, B_3)$$
$$\phi_{\sigma_4} \sim \text{Gamma}(A_4, B_4)$$
$$\phi_{\sigma_5} \sim \text{Gamma}(A_5, B_5)$$
$$s_1|\tilde{\rho} \sim \text{Categorical}_{\sigma_{1:M}}(\tilde{\rho})$$
$$s_n|s_{n-1}, \tilde{\pi}_{\sigma_{1:M}} \sim \text{Categorical}_{\sigma_{1:M}}(\pi), \qquad\qquad n = 2, \cdots, N$$
$$w_n|s_n \sim \text{Poisson}(\phi_{s_n}) \qquad\qquad n = 1, \cdots, N$$

where α, $\tilde{\beta}$, $A_{1:M}$, and $B_{1:M}$ are known parameters. We choose this model since a Poisson likelihood with Gamma prior is conjugate so we do not need to include a Metropolis–Hasting sampler inside of our Gibb's sampler.

4. Finally, we describe a MCMC method we could use to sample from this model's posterior. We describe a Gibb's Sampler and derive the conditionals which we could sample from. We begin by choosing an initial value for each of our model parameters; we use the notation ξ^0. Now we need to pick $\tilde{\rho}^1$.

$$p(\tilde{\rho}^1|\tilde{\pi}^0_{\sigma_{1:M}}, \phi^0_{\sigma_{1:M}}, s^0_{1:N}, w_{1:N}) = p(\tilde{\rho}^1|s^0_1) = \text{Dirichlet}_{\sigma_{1:M}}(\tilde{\rho}; \alpha\tilde{\beta} + \tilde{c}(s^1_1)).$$

We then sample $\tilde{\rho}^1$ from this distribution. Next we need to choose each $\tilde{\pi}^1_{\sigma_m}$ for $m = 1, \cdots, M$.

$$p(\tilde{\pi}^1_{\sigma_m}|\tilde{\rho}^1, \tilde{\pi}^0_{\sigma_{-m}}, \phi^0_{\sigma_{1:M}}, s^0_{1:N}, w_{1:N}) = p(\tilde{\pi}^1_{\sigma_m}|s^0_{1:N}) = \text{Dirichlet}_{\sigma_{1:M}}(\tilde{\pi}^1_{\sigma_m}; \alpha\tilde{\beta} + \tilde{c}(s^0)).$$

We them sample each $\tilde{\pi}_{\sigma_m}$ from this distribution. Next, we need to sample each $\phi^1_{\sigma_m}$.

$$p(\tilde{\phi}^1_{\sigma_m}|\tilde{\rho}, \tilde{\pi}^1_{\sigma_{1:M}}, \phi^0_{\sigma_{-m}}, s^0_{1:N}, w_{1:N}) = p(\tilde{\phi}^1_{\sigma_m}|, \phi^0_{\sigma_{-m}}, s^0_{1:N}, w_{1:N})$$
$$= \text{Gamma}(\phi^1_{\sigma_m}; A_m + \sum_{s_n = \sigma_m} w_n, B_m + c(s^0)).$$

Where $c(s^{\cdot})$ is the count of the number of data points in the category of interest. Lastly, we need to sample $s^1_{1:N}$. We could continue using our Gibb's Sampler to

sample each s_n individually, but this is inefficient, so we instead choose an entire sequence for $s_{1:N}$ using ancestral sampling. This is simple since each s_n relies only on s_{n-1} and $\tilde{\pi}_{\sigma_{1:M}}$. Once we have completed sampling all of ξ^1, we repeat this process J times. J is chosen to be large enough to adequately sample our posterior.