Задание априорных ограничений на все сразу

1 Честная Gate Function

$$
\log p\big(\mathbf{y}, \mathbf{Z}, \mathbf{W} | \mathbf{X}, \mathbf{V}, \mathbf{A}, \mathbf{W}^0, \boldsymbol{\Xi}, \beta, \boldsymbol{\mu}\big) =
$$

$$
= \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \left[\log \pi_k\big(\mathbf{x}_i, \mathbf{V}\big) - \frac{\beta}{2}\left(y_i - \mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i\right)^2 + \frac{1}{2}\log\frac{\beta}{2\pi}\right] +
$$

$$
+ \sum_{k=1}^{K}\left[-\frac{1}{2}\left(\mathbf{w}_k - \mathbf{w}_k^0\right)^{\mathsf{T}} \mathbf{A}_k^{-1}\left(\mathbf{w}_k - \mathbf{w}_k^0\right) + \frac{1}{2}\log\det\mathbf{A}_k^{-1} - \frac{n}{2}\log 2\pi\right] +
$$

$$
+ R\big(\mathbf{W}^0\big) + \sum_{i=1}^{N}\sum_{k=1}^{K}\mu_k \log \pi_k\big(\mathbf{x}_i, \mathbf{V}\big).
$$

1. red: правдоподобие выборки;

2. blue: априорное распределение на параметры модели (говорим что параметры на самом деле из нормального распределения с параметрами w_k^0, A_k);

3. cyan: ограничения на prior параметров модели (говорим что априорное распределение какое-то хорошее: например сумма всех w_k^0 должна равняться нулю);

4. magenta: ограничения на gate function. Конкретно в этом примере говорим, что ответы нейросети в априори должны иметь распределение Дирихле (посути вводим ограничение, что все модели не зависимо от объекта имеют некоторую вероятность);

2 Sparse Gate Function

Большой минус в работе Gate Function это то что там нету аналитической формулы для решения, поэтому приходится использовать градиентные методы которые не всегда сходятся с глобальному минимуму (это еще поверх того, что сам EM алгоритм тоже не всегда сходится к глобальному минимум, вот на этом всем у нас плохое качество получается)

Замечание. Gate function это некоторая функция, которая нужна нам только в некоторых точках. Давайте вместо востановления всей функции рассмотрим только матрицу: значение Gate Function в данных точках. Обозначим данную матрицу:

$$\hat{\boldsymbol{\pi}} = \left[\pi_k^i\right]_{k=1,i=1}^{K,N},$$

назовем $\hat{\boldsymbol{\pi}}$ как *Sparse Gate Function*.

В этом случае запишем функцию правдоподобие:

$$\begin{aligned}\log p\big(\mathbf{y}, \mathbf{Z}, \mathbf{W} | \mathbf{X}, \mathbf{A}, \mathbf{W}^0, \boldsymbol{\Xi}, \beta, \boldsymbol{\mu}\big) = &\\ &= \sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik}\left[\log \pi_k^i - \frac{\beta}{2}\left(y_i - \mathbf{w}_k^{\mathsf{T}}\mathbf{x}_i\right)^2 + \frac{1}{2}\log\frac{\beta}{2\pi}\right] + \\ &+ \sum_{k=1}^{K}\left[-\frac{1}{2}\left(\mathbf{w}_k - \mathbf{w}_k^0\right)^{\mathsf{T}}\mathbf{A}_k^{-1}\left(\mathbf{w}_k - \mathbf{w}_k^0\right) + \frac{1}{2}\log\det\mathbf{A}_k^{-1} - \frac{n}{2}\log 2\pi\right] + \\ &+ \sum_{i=1}^{N}\log\mathrm{Dir}\big(\boldsymbol{\pi}^i|\boldsymbol{\mu}, \gamma\big)\end{aligned}$$