

## PBPの学習の流れ

### パラメータの初期化

各ファイルの\_\_init\_\_()

P.130 1-8行目。  
Prior.pyで各種変数初期化。  
ただし、wと $\tilde{\{f\}}$ に関して、  
Supplementary material 6の(33),(34)で  
初期化。実用上はnatがつくもので初期化。

### 実際の学習

Pbp.py do\_pbp()

do\_first\_pass()

refine\_prior()

一定回数学習

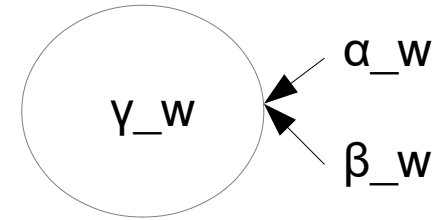
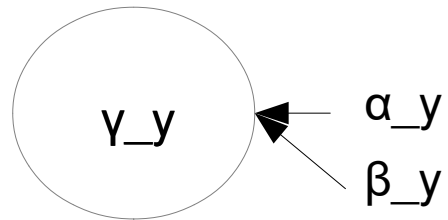
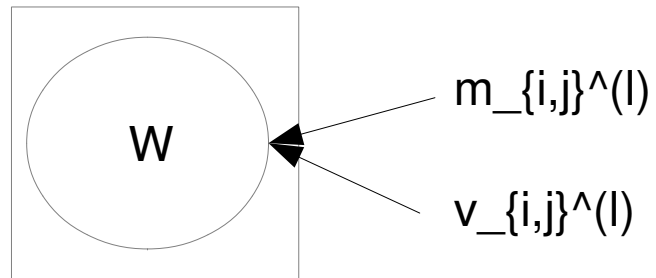
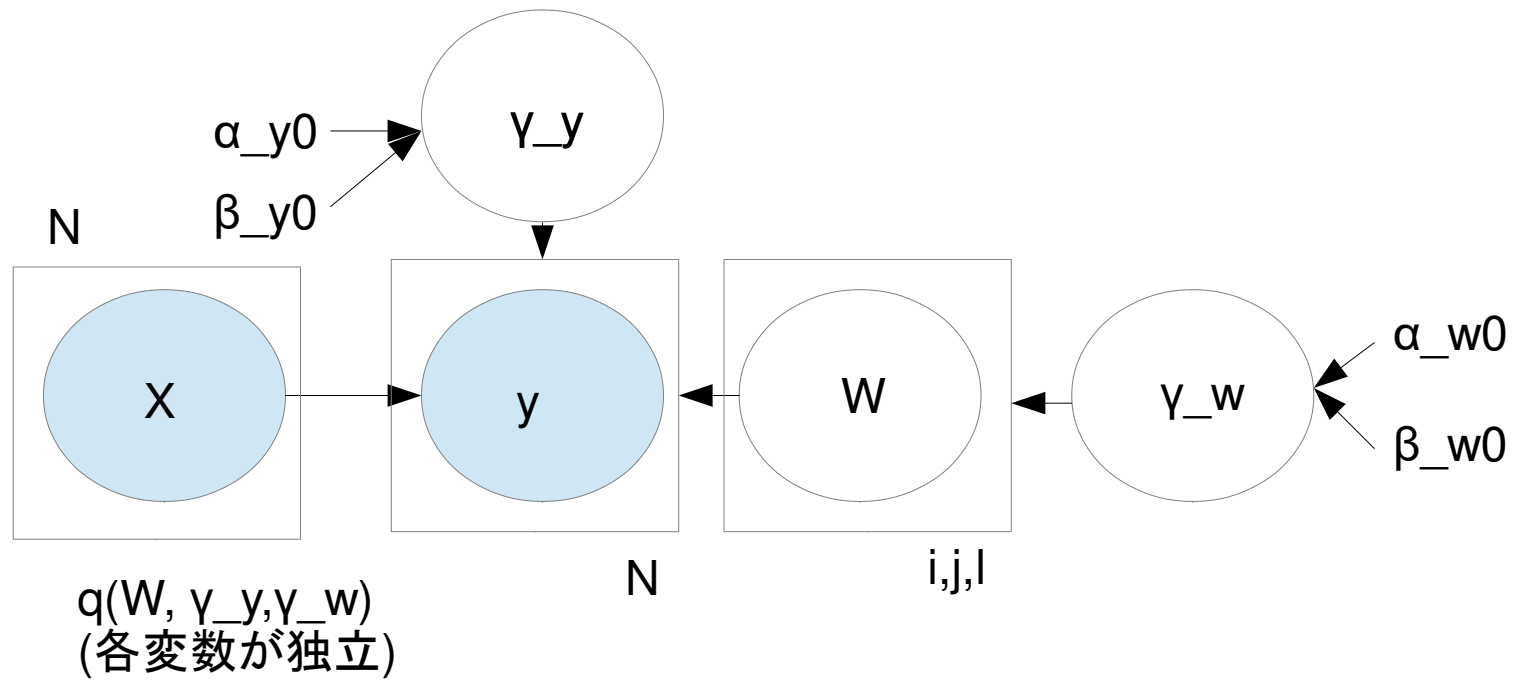
W,  $\gamma_y$ のパラメータの学習  
各教師データごとにパラメータを更新  
1, 順伝播で尤度を計算。  
計算式はpbp.pyの\_\_init\_\_で設定。  
(Network.pyのlogZ\_Z1\_Z2で(5.65)が、  
output\_probabilisticと呼ばれる、  
network\_layer.pyのoutput\_probabilisticで  
(5.66)-(5.73)を利用している。  
2, それを微分。  
Pbp.pyのadf\_updateで微分、更新。  
(実際はnewwork.pyのgenerate\_updates)  
微分はtheanoを使ってやっているので  
逆伝播がわかりにくい。  
おそらく、c版では  
supplementary material 1の(1)-(22)を  
利用していると思われる。  
(21),(22)が逆伝播で求めたい微分。  
(21),(22)の $\delta$ が逆伝搬している情報。

W,  $\gamma_w$ のパラメータの学習  
(正則化のようなことを行っているように思う。)  
5.2.5.3の8行目から。  
実際はEP法を使っている。本では  
インデックスがないことに注意。  
Supplementary materialの(35)-(39)参照。  
途中にあるように、その間で(5.59)(5.60)  
(5.61)(5.62)を利用する。

各パラメータW,  $\gamma_y$ ,  $\gamma_w$ の平均と分散が求まれば、入力に対して、出力の平均、分散が求められる  
(test\_PBP\_new.py)。検証データで平均を求めて、2乗誤差、尤度を評価。

# PBPの学習の流れ

$$P(Y|X,W,\gamma_y,\gamma_w)$$



## 4章

## 5章

| EP法  | ADF   | $W, \gamma_y$ の学習  | $W, \gamma_w$ の学習  |
|--|---|--|--|
| $\theta$<br><br>$f_n(\theta)$<br>$(n \geq 0)$<br><br>これに対して、 $p(\theta) = \prod f_n(\theta)$ | $\theta$<br><br>尤度として<br>書けるとき<br>$f_0(\theta) = p(\theta)$<br>$f_i(\theta) \propto p(D_i   \theta)$<br><br>これに対して、 $q(\theta) = \prod \tilde{f}_i(\theta)$ | $W, \gamma_y(, \gamma_w)$<br>$\gamma_w$ は積分して定数になる<br><br>$f_0(\theta) = q(W, \gamma_w, \gamma_y)$<br>$f_{\{W, \gamma_y\}}(W, \gamma_y) =$<br>$p(\gamma_w   W, \gamma_y)$<br><br>$\tilde{f}_W(\theta) = q(W)$<br>$= N(W)$<br>$\tilde{f}_{\gamma_y}(\theta) = q(\gamma_y)$<br>$= N(\gamma_y)$ | $W, \gamma_w(, \gamma_y)$<br>$\gamma_y$ は積分して定数になる<br><br>$f_{\gamma_w}(\theta) = p(\gamma_w)$<br>$f_{\gamma_y}(\theta) = p(\gamma_y)$<br>$f_{w_{i,j,l}}(\theta) = p(w_{i,j,l}   0, \gamma_w)$<br><br>$\tilde{f}_{\gamma_w} = q(\gamma_w)$<br>$= \text{Gam}(\gamma_w)$<br>$\tilde{f}_{\gamma_y} = q(\gamma_y)$<br>$= \text{Gam}(\gamma_y)$<br>$w_{i,j,l}$ を含む部分について、<br>$\tilde{f}_{w_{i,j,l}}(\theta)$<br>$= N(w_{i,j,l}) \text{Gam}(\gamma_w)$ |