$N\{\vec{x}_1,\dots,\vec{x}_N\}\vec{x}_i \in \mathbb{R}^D\{y_1^{(1)},\dots,y_n^{(1)},\dots,y_1^{(T)},\dots,y_N^{(T)}\}y_i^{(t)} \in \mathbb{Z}_2\vec{x}_i t\{z_1,\dots,z_N\}z_i \in \mathbb{Z}_2\vec{x}yzp(z \mid \vec{x})$

$$
\begin{aligned}
X &= [\vec{x}_1^T;\dots;\vec{x}_N^T] \in \mathbb{R}^{N\times D}\\
Y &= [y_1^{(1)},\dots,y_1^{(T)};\dots;y_N^{(1)},\dots,y_N^{(T)}] \in \mathbb{Z}_2^{N\times T}\\
Z &= (z_1,\dots,z_N) \in \mathbb{Z}_2^N
\end{aligned}
$$

$$
\begin{aligned}
p(y_i^{(t)} \mid \vec{x}_i, z_i, \vec{w}_t, \gamma_t) &= (1-\eta_t(\vec{x}_i \mid \vec{w}_t,\gamma_t))^{|y_i^{(t)}-z_i|}\eta_t(\vec{x}_i \mid \vec{w}_t,\gamma_t)^{1-|y_i^{(t)}-z_i|}\\
\eta_t(\vec{x}_i \mid \vec{w}_t,\gamma_t) &= \sigma(\vec{w}_t^T\vec{x}_i - \gamma_t)
\end{aligned}
$$

$$p(z_i = 1 \mid \vec{x}_i, \alpha, \beta) = \sigma(\vec{\alpha}^T\vec{x}_i + \beta)$$

$\vec{\theta} = \{\vec{\alpha},\beta,\vec{w}_1,\dots,\vec{w}_T,\gamma_1,\dots,\gamma_T\}$
$p(Y \mid X, \vec{\theta})$

$$
\begin{aligned}
\vec{\theta}^\star &= _{\vec{\theta}}\sum_{i=1}^N\sum_{t=1}^T p(y_i^{(t)} \mid \vec{x}_i,\vec{\theta})\\
&= _{\vec{\theta}}\sum_{i=1}^N\sum_{t=1}^T\sum_{z_i=0}^1 p(y_i^{(t)}, z_i \mid \vec{x}_i,\vec{\theta})
\end{aligned}
$$

z_i

$p(z_i \mid \vec{x}_i, y_i^{(1)},\dots,y_i^{(T)},\vec{\theta}) i = 1,\dots,N$

$$
\begin{aligned}
p(z_i \mid \vec{x}_i, y_i^{(1)},\dots,y_i^{(T)},\vec{\theta}) &= \frac{1}{A_i}p(z_i, y_i^{(1)},\dots,y_i^{(T)} \mid \vec{x}_i,\vec{\theta})\\
&= \frac{1}{A_i}\prod_{t=1}^T p(z_i, y_i^{(t)} \mid \vec{x}_i,\vec{\theta})\\
&= \frac{1}{A_i}\prod_{t=1}^T p(y_i^{(t)} \mid \vec{x}_i, z_i,\vec{\theta})p(z_i \mid \vec{x}_i,\vec{\theta})\\
&= \frac{1}{A_i}\prod_{t=1}^T p(y_i^{(t)} \mid \vec{x}_i, z_i,\vec{w}_t,\gamma_t)p(z_i \mid \vec{x}_i,\vec{\alpha},\beta)
\end{aligned}
$$

A_i

$$A_i = \sum_{z_i=0}^1 p(z_i, y_i^{(1)},\dots,y_i^{(T)} \mid \vec{x}_i,\vec{\theta}).$$

$\tilde{p}(z_i) = p(z_i \mid \vec{x}_i, y_i^{(1)},\dots,y_i^{(T)},\vec{\theta})$

$$\vec{\theta} = _{\vec{\theta}}\sum_{i=1}^N Q_i(\vec{\theta},\vec{\theta})$$

$$Q_i(\vec{\theta}, \vec{\theta}) = \sum_{z_i=0}^{1} p(z_i \mid \vec{x}_i, y_i^{(1)}, \ldots, y_i^{(T)}, \vec{\theta})\ p(\vec{x}_i, y_i^{(1)}, \ldots, y_i^{(T)}, z_i \mid \vec{\theta}).$$

$$\begin{aligned}
Q_i(\vec{\theta}, \vec{\theta}) &= \sum_{z_i=0}^{1} \tilde{p}(z_i)\ p(\vec{x}_i, y_i^{(1)}, \ldots, y_i^{(T)}, z_i \mid \vec{\theta}) \\
&= \sum_{z_i=0}^{1} \sum_{t=1}^{T} \tilde{p}(z_i)\ p(\vec{x}_i, y_i^{(t)}, z_i \mid \vec{\theta}) \\
&= \sum_{z_i=0}^{1} \sum_{t=1}^{T} \tilde{p}(z_i)\ (p(y_i^{(t)}, z_i \mid \vec{x}_i, \vec{\theta}) p(\vec{x}_i \mid \vec{\theta})) \\
&= T\ p(\vec{x}_i \mid \vec{\theta}) + \sum_{z_i=0}^{1} \sum_{t=1}^{T} \tilde{p}(z_i)\ p(y_i^{(t)}, z_i \mid \vec{x}_i, \vec{\theta}) \\
&= T\ p(\vec{x}_i \mid \vec{\theta}) + \\
&\sum_{z_i=0}^{1} \sum_{t=1}^{T} \tilde{p}(z_i)(\ p(y_i^{(t)} \mid \vec{x}_i, z_i, \vec{\theta}) +\ p(z_i \mid \vec{x}_i, \vec{\theta})) \\
&= T\ p(\vec{x}_i \mid \vec{\theta}) + \\
&\sum_{z_i=0}^{1} \sum_{t=1}^{T} \tilde{p}(z_i)(\ p(y_i^{(t)} \mid \vec{x}_i, z_i, \vec{w}_t, \gamma_t) +\ p(z_i \mid \vec{x}_i, \vec{\alpha}, \beta))
\end{aligned}$$

$$\vec{\theta} = {}_{\vec{\theta}} \sum_{i=1}^{N} \sum_{z_i=0}^{1} \sum_{t=1}^{T} \tilde{p}(z_i)(\ p(y_i^{(t)} \mid \vec{x}_i, z_i, \vec{w}_t, \gamma_t) +\ p(z_i \mid \vec{x}_i, \vec{\alpha}, \beta))$$

$T\ p(\vec{x}_i \mid \vec{\theta})\vec{\theta}x_i$

$$f(\vec{\theta}) = \sum_{i=1}^{N} \sum_{z_i=0}^{1} \sum_{t=1}^{T} \tilde{p}(z_i)(\ p(y_i^{(t)} \mid \vec{x}_i, z_i, \vec{w}_t, \gamma_t) +\ p(z_i \mid \vec{x}_i, \vec{\alpha}, \beta)).$$

$\tilde{p}(z_i)\vec{\theta}$
$f\vec{\theta}$

$\nabla_{\vec{\alpha}} f(\vec{\theta})$

$$
\begin{aligned}
\nabla_{\vec{\alpha}} f(\vec{\theta}) &= T \sum_{i=1}^{N} \sum_{z_i=0}^{1} \nabla_{\vec{\alpha}} (\tilde{p}(z_i)\ p(z_i \mid \vec{x}_i, \vec{\alpha}, \beta)) \\
&= T \sum_{i=1}^{N} \sum_{z_i=0}^{1} \tilde{p}(z_i) \nabla_{\vec{\alpha}}\ p(z_i \mid \vec{x}_i, \vec{\alpha}, \beta) \\
&= T \sum_{i=1}^{N} \tilde{p}(z_i = 1) \nabla_{\vec{\alpha}}\ \sigma(\vec{\alpha}^T \vec{x}_i + \beta) + \tilde{p}(z_i = 0) \nabla_{\vec{\alpha}} (1 -\ \sigma(\vec{\alpha}^T \vec{x}_i + \beta)) \\
&= T \sum_{i=1}^{N} (\tilde{p}(z_i = 1) - \tilde{p}(z_i = 0)) \nabla_{\vec{\alpha}}\ \sigma(\vec{\alpha}^T \vec{x}_i + \beta) \\
&= T \sum_{i=1}^{N} (\tilde{p}(z_i = 1) - \tilde{p}(z_i = 0)) \frac{\nabla_{\vec{\alpha}} \sigma(\vec{\alpha}^T \vec{x}_i + \beta)}{\sigma(\vec{\alpha}^T \vec{x}_i + \beta)} \\
&= T \sum_{i=1}^{N} (\tilde{p}(z_i = 1) - \tilde{p}(z_i = 0)) \frac{\sigma(\vec{\alpha}^T \vec{x}_i + \beta)(1 - \sigma(\vec{\alpha}^T \vec{x}_i + \beta))}{\sigma(\vec{\alpha}^T \vec{x}_i + \beta)} \vec{x} \\
&= T \sum_{i=1}^{N} (\tilde{p}(z_i = 1) - \tilde{p}(z_i = 0))(1 - \sigma(\vec{\alpha}^T \vec{x}_i + \beta)) \vec{x}
\end{aligned}
$$

$\frac{\partial f}{\partial \beta}(\vec{\theta}) \nabla_{\vec{\alpha}} f(\vec{\theta}) \frac{\partial}{\partial \beta}(\vec{\alpha}^T \vec{x}_i + \beta) = 1$

$$
\frac{\partial f}{\partial \beta}(\vec{\theta}) = T \sum_{i=1}^{N} (\tilde{p}(z_i = 1) - \tilde{p}(z_i = 0))(1 - \sigma(\vec{\alpha}^T \vec{x}_i + \beta))
$$

$\nabla_{\vec{w}_t} f(\vec{\theta})$

$$
\begin{aligned}
\nabla_{\vec{w}_t} f(\vec{\theta}) &= \sum_{i=1}^{N} \sum_{z_i=0}^{1} \sum_{s=1}^{T} \nabla_{\vec{w}_t} (\tilde{p}(z_i)\ p(y_i^{(s)} \mid \vec{x}_i, z_i, \vec{w}_s, \gamma_s)) \\
&= \sum_{i=1}^{N} \sum_{z_i=0}^{1} \nabla_{\vec{w}_t} (\tilde{p}(z_i)\ p(y_i^{(t)} \mid \vec{x}_i, z_i, \vec{w}_t, \gamma_t)) \\
&= \sum_{i=1}^{N} \sum_{z_i=0}^{1} \tilde{p}(z_i) \nabla_{\vec{w}_t}\ p(y_i^{(t)} \mid \vec{x}_i, z_i, \vec{w}_t, \gamma_t) \\
&= \sum_{i=1}^{N} \sum_{z_i=0}^{1} \tilde{p}(z_i) \nabla_{\vec{w}_t} ((1 - \eta_t(\vec{x}_i \mid \vec{w}_t, \gamma_t))^{|y_i^{(t)} - z_i|} \eta_t(\vec{x}_i \mid \vec{w}_t, \gamma_t)^{1 - |y_i^{(t)} - z_i|}) \\
&= \sum_{i=1}^{N} \sum_{z_i=0}^{1} \tilde{p}(z_i) (\nabla_{\vec{w}_t} (1 - \eta_t(\vec{x}_i \mid \vec{w}_t, \gamma_t))^{|y_i^{(t)} - z_i|} \eta_t(\vec{x}_i \mid \vec{w}_t, \gamma_t)^{1 - |y_i^{(t)} - z_i|} + \nabla_{\vec{w}_t} (1 - \eta_t(\vec{x}_i \mid \vec{w}_t, \gamma_t))^{|y_i^{(t)} - z_i|} \eta_t(\vec{x}_i \mid
\end{aligned}
$$