

# MACHINE LEARNING 1: ASSIGNMENT 7

Linxi Wang 587032  
Lusine Nazaretyan 513624  
Thomas Herold 535025  
Ya Qian 518902  
Karen Nazaretyan

## 1 Discrete EM, coin tosses from multiple distributions

Since we want to find the parameter  $\theta = (\lambda, p_1, p_2)$  that maximizes

$$Q(\theta, \theta^{old}) = \sum_{z \in \{heads, tails\}^N} \underbrace{P(Z = z | X = x, \theta^{old})}_{=q(z)} \underbrace{\log(P(X = x, Z = z | \theta))}_{=F(z, \theta)},$$

The factor  $P(Z = z | X = x, \theta^{old})$  does not depend on  $\theta$  and will therefore remain unchanged under differentiation.

Therefore, we only have to care about the function  $F(z, \theta)$  for now.

Let:

$H(v) \equiv$  number of heads in  $v$

$T(v) \equiv$  number of tails in  $v$

for any vector  $v$  (including scalars).

We can obtain:

$$\begin{aligned} F(z, \theta) &= \log(P(X = x, Z = z | \theta)) \\ &= \log\left(\prod_{i=1}^N P(Z = z^{(i)} | \theta) \prod_{j=1}^m P(X_j = x_j^{(i)} | Z = z^{(i)}, \theta)\right) \\ &= \sum_{i=1}^N \log(P(Z = z^{(i)} | \theta)) + \sum_{j=1}^m \log(P(X_j = x_j^{(i)} | Z = z^{(i)}, \theta)). \end{aligned}$$

With our prior knowledge, we can now rewrite

$$\begin{aligned} P(Z = z^{(i)} | \theta) &= \lambda^{H(z^{(i)})} (1 - \lambda)^{T(z^{(i)})} \\ \text{and } P(X_j = x_j^{(i)} | Z = z^{(i)}, \theta) &= (p_1^{H(x_j^{(i)})} (1 - p_1)^{T(x_j^{(i)})})^{H(z^{(i)})} (p_2^{H(x_j^{(i)})} (1 - p_2)^{T(x_j^{(i)})})^{T(z^{(i)})}. \end{aligned}$$

If we insert this into the equation above, we get:

$$\begin{aligned} F(z, \theta) &= \sum_{i=1}^N \log(\lambda^{H(z^{(i)})} (1 - \lambda)^{T(z^{(i)})}) \\ &\quad + \sum_{j=1}^m \log\left((p_1^{H(x_j^{(i)})} (1 - p_1)^{T(x_j^{(i)})})^{H(z^{(i)})} (p_2^{H(x_j^{(i)})} (1 - p_2)^{T(x_j^{(i)})})^{T(z^{(i)})}\right) \\ &= \sum_{i=1}^N H(z^{(i)}) \log(\lambda) + T(z^{(i)}) \log(1 - \lambda) + \sum_{j=1}^m H(x_j^{(i)}) H(z^{(i)}) \log(p_1) \\ &\quad + T(x_j^{(i)}) H(z^{(i)}) \log(1 - p_1) + H(x_j^{(i)}) T(z^{(i)}) \log(p_2) + T(x_j^{(i)}) T(z^{(i)}) \log(1 - p_2). \end{aligned}$$

### Parameter $\hat{\lambda}$

We now want to derive an expression for  $\hat{\lambda}$  and therefore we set the derivative of  $Q$  w.r.t.  $\lambda$  to zero:

$$\frac{dQ(\theta, \theta_{old})}{d\lambda} = \sum_{z \in \{T, H\}^N} q(z) \frac{dF(z, \theta)}{d\lambda} \stackrel{!}{=} 0. \quad (1)$$

So we see that we have to compute the derivative of  $F(z, \theta)$  w.r.t.  $\lambda$ .

$$\begin{aligned}
\frac{dF(z, \theta)}{d\lambda} &= \frac{d}{d\lambda} \sum_{i=1}^N H(z^{(i)}) \log(\lambda) + T(z^{(i)}) \log(1 - \lambda) \\
&= \frac{d}{d\lambda} \log(\lambda) \sum_{i=1}^N H(z^{(i)}) + \frac{d}{d\lambda} \log(1 - \lambda) \sum_{i=1}^N T(z^{(i)}) \\
&= \frac{H(z)}{\lambda} - \frac{T(z)}{1 - \lambda} \\
&= \frac{(1 - \lambda)H(z) - \lambda T(z)}{\lambda(1 - \lambda)} \\
&\stackrel{T(z)=N-H(z)}{=} \frac{H(z) - \lambda N}{\lambda(1 - \lambda)}
\end{aligned}$$

We can now insert this result into equation (1):

$$\begin{aligned}
&\sum_{z \in \{T, H\}^N} q(z) \frac{dF(z, \theta)}{d\lambda} = 0 \\
\Leftrightarrow &\sum_{z \in \{T, H\}^N} q(z) \frac{H(z) - \lambda N}{\lambda(1 - \lambda)} = 0 \\
\Leftrightarrow &\sum_{z \in \{T, H\}^N} q(z) \frac{H(z)}{\lambda(1 - \lambda)} = \sum_{z \in \{T, H\}^N} q(z) \frac{\lambda N}{\lambda(1 - \lambda)} \\
\Leftrightarrow &\sum_{z \in \{T, H\}^N} q(z) H(z) = \lambda N \sum_{z \in \{T, H\}^N} q(z) \\
\Leftrightarrow &\frac{\sum_{z \in \{T, H\}^N} q(z) H(z)}{N \sum_{z \in \{T, H\}^N} q(z)} = \lambda
\end{aligned}$$

Since  $q(z)$  represents a probability distribution, we know that  $\sum_{z \in \{T, H\}^N} q(z) = 1$  and therefore we obtain

$$\hat{\lambda} = \frac{\sum_{z \in \{T, H\}^N} q(z) H(z)}{N}.$$

## Parameter $\hat{p}_1$

We now want to derive an expression for  $\hat{p}_1$  and therefore we set the derivative of  $Q$  w.r.t.  $p_1$  to zero:

$$\frac{dQ(\theta, \theta_{old})}{dp_1} = \sum_{z \in \{T, H\}^N} q(z) \frac{dF(z, \theta)}{dp_1} \stackrel{!}{=} 0. \quad (2)$$

So we see that we have to compute the derivative of  $F(z, \theta)$  w.r.t.  $p_1$ .

$$\begin{aligned}
\frac{dF(z, \theta)}{dp_1} &= \frac{d}{dp_1} \left( \sum_{i=1}^N \sum_{j=1}^m H(x_j^{(i)}) H(z^{(i)}) \log(p_1) + T(x_j^{(i)}) H(z^{(i)}) \log(1 - p_1) \right) \\
&= \frac{d}{dp_1} \log(p_1) \sum_{i=1}^N H(z^{(i)}) \sum_{j=1}^m H(x_j^{(i)}) + \frac{d}{dp_1} \log(1 - p_1) \sum_{i=1}^N H(z^{(i)}) \sum_{j=1}^m T(x_j^{(i)}) \\
&= \frac{1}{p_1} \sum_{i=1}^N H(z^{(i)}) H(x^{(i)}) - \frac{1}{1 - p_1} \sum_{i=1}^N H(z^{(i)}) T(x^{(i)}) \\
&\stackrel{T(x^{(i)}) = m - H(x^{(i)})}{=} \frac{1}{p_1} \sum_{i=1}^N H(z^{(i)}) H(x^{(i)}) - \frac{1}{1 - p_1} \sum_{i=1}^N m H(z^{(i)}) + \frac{1}{1 - p_1} \sum_{i=1}^N H(x^{(i)}) H(z^{(i)}) \\
&= \frac{1}{p_1(1 - p_1)} \sum_{i=1}^N H(z^{(i)}) H(x^{(i)}) - \frac{m}{1 - p_1} H(z).
\end{aligned}$$

We can now insert this expression into equation (2):

$$\begin{aligned}
&\sum_{z \in \{T, H\}^N} q(z) \frac{dF(z, \theta)}{dp_1} = 0 \\
\Leftrightarrow \sum_{z \in \{T, H\}^N} q(z) \left( \frac{1}{p_1(1 - p_1)} \sum_{i=1}^N H(z^{(i)}) H(x^{(i)}) - \frac{m}{1 - p_1} H(z) \right) &= 0 \\
\Leftrightarrow \sum_{z \in \{T, H\}^N} q(z) \frac{1}{p_1(1 - p_1)} \sum_{i=1}^N H(z^{(i)}) H(x^{(i)}) &= \sum_{z \in \{T, H\}^N} q(z) \frac{m}{1 - p_1} H(z) \\
\Leftrightarrow \sum_{z \in \{T, H\}^N} q(z) \sum_{i=1}^N H(z^{(i)}) H(x^{(i)}) &= m \cdot p_1 \sum_{z \in \{T, H\}^N} q(z) H(z)
\end{aligned}$$

and therefore we get:

$$\begin{aligned}
\hat{p}_1 &= \frac{\sum_{z \in \{T, H\}^N} q(z) \sum_{i=1}^N H(z^{(i)}) H(x^{(i)})}{m \sum_{z \in \{T, H\}^N} q(z) H(z)} \\
&= \frac{\sum_{z \in \{T, H\}^N} q(z) \sum_{i=1}^N H(z^{(i)}) H(x^{(i)})}{m \cdot N \cdot \hat{\lambda}}.
\end{aligned}$$

## Parameter $\hat{p}_2$

In order to determine an explicit expression for  $\hat{p}_2$ , we proceed analogously to the case of  $\hat{p}_1$ . We set the derivative of  $Q$  w.r.t.  $p_2$  to zero:

$$\frac{dQ(\theta, \theta_{old})}{dp_2} = \sum_{z \in \{T, H\}^N} q(z) \frac{dF(z, \theta)}{dp_2} \stackrel{!}{=} 0 \tag{3}$$

and compute the derivative of  $F(z, \theta)$  w.r.t.  $p_2$ :

$$\begin{aligned}
\frac{dF(z, \theta)}{dp_2} &= \frac{d}{dp_2} \left( \sum_{i=1}^N \sum_{j=1}^m H(x_j^{(i)}) T(z^{(i)}) \log(p_2) + T(x_j^{(i)}) T(z^{(i)}) \log(1 - p_2) \right) \\
&= \frac{d}{dp_2} \log(p_2) \sum_{i=1}^N \sum_{j=1}^m H(x_j^{(i)}) T(z^{(i)}) + \frac{d}{dp_2} \log(1 - p_2) \sum_{i=1}^N \sum_{j=1}^m T(x_j^{(i)}) T(z^{(i)}) \\
&= \frac{1}{p_2} \sum_{i=1}^N H(x^{(i)}) T(z^{(i)}) - \frac{1}{1 - p_2} \sum_{i=1}^N T(x^{(i)}) T(z^{(i)}) \\
&\stackrel{T(x^{(i)}) = m - H(x^{(i)})}{=} \frac{1}{p_2} \sum_{i=1}^N H(x^{(i)}) T(z^{(i)}) - \frac{1}{1 - p_2} \sum_{i=1}^N m T(z^{(i)}) + \frac{1}{1 - p_2} \sum_{i=1}^N H(x^{(i)}) T(z^{(i)}) \\
&= \frac{1}{p_2(1 - p_2)} \sum_{i=1}^N H(x^{(i)}) T(z^{(i)}) - \frac{m}{1 - p_2} T(z).
\end{aligned}$$

We can now insert this into equation (3) and obtain:

$$\begin{aligned}
&\sum_{z \in \{T, H\}^N} q(z) \frac{dF(z, \theta)}{dp_2} = 0 \\
\Leftrightarrow \sum_{z \in \{T, H\}^N} q(z) \left( \frac{1}{p_2(1 - p_2)} \sum_{i=1}^N H(x^{(i)}) T(z^{(i)}) - \frac{m}{1 - p_2} T(z) \right) &= 0 \\
\Leftrightarrow \sum_{z \in \{T, H\}^N} q(z) \frac{1}{p_2(1 - p_2)} \sum_{i=1}^N H(x^{(i)}) T(z^{(i)}) &= \sum_{z \in \{T, H\}^N} q(z) \frac{m}{1 - p_2} T(z) \\
\Leftrightarrow \sum_{z \in \{T, H\}^N} q(z) \sum_{i=1}^N H(x^{(i)}) T(z^{(i)}) &= m \cdot p_2 \sum_{z \in \{T, H\}^N} q(z) T(z).
\end{aligned}$$

This yields:

$$\hat{p}_2 = \frac{\sum_{z \in \{T, H\}^N} q(z) \sum_{i=1}^N H(x^{(i)}) T(z^{(i)})}{m \sum_{z \in \{T, H\}^N} q(z) T(z)}$$