Yang, Ya-Chu
11571276
qwerthjkl45@gmail.com

Homework Assignment 3
Machine Learning 1, 17/18

2017-09-25

1 Naive Bayes Spam Classification

1.

$$
\begin{aligned}
p(\boldsymbol{T},\boldsymbol{X}|\boldsymbol{\theta}) &= \prod_{n=1}^{N}[\prod_{k=1}^{3}[p(C_k)p(x_n|C_k)]^{\mathbb{I}(t_n=k)}] \\
&= \prod_{n=1}^{N}[\prod_{k=1}^{3}[p(C_k)[\prod_{d=1}^{D}p(x_{dn}|C_k,\theta_{dk}))]]^{\mathbb{I}(t_n=k)}] \\
&= \prod_{n=1}^{N}[\prod_{k=1}^{3}[\pi_k[\prod_{d=1}^{D}p(x_{dn}|C_k,\theta_{dk}))]]^{\mathbb{I}(t_n=k)}]
\end{aligned}
$$

2.

$$
\begin{aligned}
p(\boldsymbol{T},\boldsymbol{X}|\boldsymbol{\theta}) &= \prod_{n=1}^{N}[\prod_{k=1}^{3}[\pi_k[\prod_{d=1}^{D}p(x_{dn}|C_k,\theta_{dk}))]]^{\mathbb{I}(t_n=k)}] \\
&= \prod_{n=1}^{N}[\prod_{k=1}^{3}[\pi_k[\prod_{d=1}^{D}\frac{\lambda_{dk}^{x_{nd}}}{x_{nd}!}exp(-\lambda_{dk})]]^{\mathbb{I}(t_n=k)}]
\end{aligned}
$$

3.

$$
log(p(\boldsymbol{T},\boldsymbol{X}|\boldsymbol{\theta}) = \sum_{n=1}^{N}\sum_{k=1}^{3}[\mathbb{I}(t_n=k)[log(\pi_k)+\sum_{d=1}^{D}[x_{nd}log(\lambda_{dk}-log(x_d!)-\lambda_{dk})]]]
$$

4.

$$
\begin{aligned}
&\partial[\frac{log(p(\boldsymbol{T},\boldsymbol{X}|\boldsymbol{\theta})}{\partial\lambda_{dk}}] = \sum_{n=1}^{N}\mathbb{I}(t_n=k)[\frac{x_{nd}}{\lambda_{dk}}-1]=0 \\
&\therefore \sum_{n=1}^{N}\mathbb{I}(t_n=k)[\frac{x_{nd}}{\lambda_{dk}}] = \sum_{n=1}^{N}\mathbb{I}(t_n=k)[1] \\
&\therefore \frac{1}{\lambda_{dk}}\sum_{n=1}^{N}\mathbb{I}(t_n=k)x_{nd} = \sum_{n=1}^{N}\mathbb{I}(t_n=k)[1] \\
&\therefore \lambda_{dk} = \sum_{n=1}^{N}\mathbb{I}(t_n=k)x_{nd}
\end{aligned}
$$

5.

$$
p(C_1|x) = \frac{p(x|C_1)p(C_1)}{p(x|C_1)p(C_1)+p(x|C_2)p(C_2)+p(x|C_3)p(C_3)}
$$

6.

$$
p(C_1|x) = \frac{\pi_1\prod_{d=1}^{D}\frac{\lambda_{d1}^{x_{nd}}}{x_{nd}!}exp(-\lambda_{d1})}{\sum_{k=1}^{3}[\pi_k\prod_{d=1}^{D}[\frac{\lambda_{d1}^{x_{nd}}}{x_{nd}!}exp(-\lambda_{d1})]]}
$$

7.

$$p(C_1|x) = \frac{exp(a_1)}{\sum_j exp(a_j)}$$

$$\therefore a_1 = ln[p(x|C_1)p(C_1)] = ln[\pi_1 \prod_{d=1}^{D} \frac{\lambda_{d1}^{x_{nd}}}{x_{nd}!} exp(-\lambda_{d1})]$$

$$= ln\pi_1 + \sum_{d=1}^{D} [x_{nd} ln(\lambda_{d_1}) - ln(x_{nd}!) - \lambda_{d1}]$$

2 Multi-class Logistic Regression

1.

$$\frac{\partial y_k}{\partial w_j} = \frac{\partial y_k}{\partial a_j} \frac{\partial a_j}{\partial w_j} = \frac{\partial \frac{exp(a_k)}{\sum_i exp(a_i)}}{\partial a_j} \frac{\partial w^T \phi}{\partial w_j}$$

$$= \frac{\boldsymbol{I}_{kj} exp(a_k)[\sum_i exp(a_i)] - exp(a_k)exp(a_j)}{(\sum_i exp(a_i))^2} \phi$$

$$= (\boldsymbol{I} y_k(\phi_n) - y_k(\phi_n) y_j(\phi_n)\phi_n$$

$$= y_k(\phi_n)(\boldsymbol{I} - y_j(\phi_n))\phi_n$$

2.

$$p(T|W) = \prod_{n=1}^{N} \prod_{k=1}^{K} p(C_k|\phi_n)^{t_{nk}} = \prod_{n=1}^{N} \prod_{k=1}^{K} y_k(\phi_n)^{t_{nk}}$$

$$logp(T|W) = \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} log(y_k(\phi_n))$$

3.

$$\frac{\partial log(p(T|W)}{\partial w_j} = \sum_{n=1}^{N} \sum_{k=1}^{K} \frac{t_{nk}}{y_k(\phi_n)} \frac{\partial y_k(\phi_n)}{\partial w_j} = \sum_{n=1}^{N} \sum_{k=1}^{K} \frac{t_{nk}}{y_k(\phi_n)} y_k(\phi_n)(\boldsymbol{I} - y_j(\phi_n))\phi_n$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} [t_{nk}(\boldsymbol{I_{kj}} - y_j(\phi_n))\phi_n] = \sum_{n=1}^{N} \sum_{k=1}^{K} [t_{nj} - t_{nk} y_j(\phi_n)]\phi_n$$

$$= \sum_{n=1}^{N} [t_{nj} - y_j(\phi_n)]\phi_n$$

4.

$$\frac{\partial E(W)}{\partial w_j} = \frac{\partial(-log(p(T|W))}{\partial w_j} = \sum_{n=1}^{N} [y_j(\phi_n) - t_{nj}]\phi_n$$