Tian Lu
tianlu@g.harvard.edu
Due: Collaborators:

Problem 1 (Naive Bayes problem sets)

In this problem, we look at maximum likelihood parameter estimation using the naive Bayes assumption. Here, the input features $x_j, j = 1, ..., n$ to our model are discrete, binary-valued variables, so $x_j \in \{0, 1\}$. We call $x = [x_1, x_2, \ldots x_n]^T$ to be the input vector. For each training example, our output targets are a single binary-value $y0, 1$. Our model is then parameterized by $\theta_{j|y=0} = p(x_j = 1|y = 0), \theta_{j|y=1} = p(x_j = 1|y = 1)$, and $\theta_y = p(y = 1)$. We model the joint distribution of (x, y) according to

$$P(y) = (\theta_y)^y (1 - \theta_y)^{1-y}$$
$$p(x|y = 0) = \prod_{j=1}^{n} p(x_j|y = 0)$$
$$p(x|y = 1) = \prod_{j=1}^{n} p(x_j|y = 1)$$

(a) Find the joint likelihood function $L = log \prod_{i=1}^{m} p(x^{(i)}, y^{(i)}; \theta)$ in terms of the model parameters given above. Here, θ represents the entire set of parameters$\{\theta_y, \theta_{j|y=0}, \theta_{j|y=1}, j = 1, \ldots, n\}$.

(b) Derive the parameters $\theta_y, \theta_{j|y=0}, \theta_{j|y=1}$ which maximize the likelihood function

(a)

$$\begin{aligned}
L &= log \prod_{i=1}^{m} p(x^{(i)}, y^{(i)}; \theta) \\
&= log \prod_{i=1}^{m} p(x^{(i)}|y^{(i)}; \theta) P(y^{(i)}; \theta) \\
&= log \prod_{i=1}^{m} (\prod_{j=1}^{n} p(x_j^{(i)}|y^{(i)}; \theta)) P(y^{(i)}; \theta) \\
&= \sum_{i=1}^{m} (\sum_{j=1}^{n} logp(x_j^{(i)}|y^{(i)}; \theta) + logP(y^{(i)}; \theta) \\
&= \sum_{i=1}^{m} (\sum_{j=1}^{n} (x_j^{(i)} log\theta_{j|y^{(i)}} + (1 - x_j^{(i)}) log(1 - \theta_{j|y^{(i)}}) + y^{(i)} log\theta_y + (1 - y^{(i)} log(1 - \theta_y))
\end{aligned}$$

(b)

$$
\begin{aligned}
\nabla_{\theta_{j|y=0}} L &= \nabla_{\theta_{j|y=0}} \sum_{j=1}^{m} (x_j^{(i)} log\theta_{j|y^{(i)}} + (1-x_j^{(i)}) log(1-\theta_{j|y^{(i)}})\\
&= \nabla_{\theta_{j|y=0}} \sum_{j=1}^{m} (x_j^{(i)} log\theta_{j|y=0} 1\{y^{(i)}=0\} + (1-x_j^{(i)}) log(1-\theta_{j|y=0}) 1\{y^{(i)}=0\})\\
&= \sum_{j=1}^{m} (\frac{x_j^{(i)} 1\{y^{(i)}=0\}}{\theta_{j|y=0}} + \frac{(1-x_j^{(i)}) 1\{y^{(i)}=0\}}{1-\theta_{j|y=0}})\\
0 &= \sum_{j=1}^{m} (\frac{x_j^{(i)} 1\{y^{(i)}=0\}}{\theta_{j|y=0}} + \frac{(1-x_j^{(i)}) 1\{y^{(i)}=0\}}{1-\theta_{j|y=0}})\\
&= \sum_{j=1}^{m} (x_j^{(i)}(1-\theta_{j|y=0}) 1\{y^{(i)}=0\} + (1-x_j^{(i)})\theta_{j|y=0} 1\{y^{(i)}=0\})\\
&= \sum_{j=1}^{m} ((x_j^{(i)}-\theta_{j|y=0}) 1\{y^{(i)}=0\})\\
&= \sum_{j=1}^{m} (x_j^{(i)} 1\{y^{(i)}=0\}) - \sum_{j=1}^{m} (\theta_{j|y=0}) 1\{y^{(i)}=0\}\\
&= \sum_{j=1}^{m} (1\{x_j^{(i)}=1 \wedge y^{(i)}=0\}) - \theta_{j|y=0} \sum_{j=1}^{m} 1\{y^{(i)}=0\}\\
\theta_{j|y=0} &= \frac{\sum_{j=1}^{m} (1\{x_j^{(i)}=1 \wedge y^{(i)}=0\})}{\sum_{j=1}^{m} 1\{y^{(i)}=0\}}
\end{aligned}
$$

$Therefore,$

$$
\begin{aligned}
\theta_{j|y=1} &= \frac{\sum_{j=1}^{m} (1\{x_j^{(i)}=1 \wedge y^{(i)}=1\})}{\sum_{j=1}^{m} 1\{y^{(i)}=1\}}\\
\nabla_{\theta_y} L &= \nabla_{\theta_y} \sum_{j=1}^{m} (y^{(i)} log\theta_y + (1-y^{(i)} log(1-\theta_y))\\
0 &= \sum_{j=1}^{m} (y^{(i)}(1-\theta_y) + (1-y^{(i)}\theta_y)\\
&= \sum_{j=1}^{m} y^{(i)} - \sum_{j=1}^{m} \theta_y\\
\theta_y &= \frac{\sum_{j=1}^{m} 1\{y^{(i)}=1\}}{m}
\end{aligned}
$$