HW02

problem 1:

(a)
$$L = \sum_{i=1}^{n} \ln \pi^{y_i}(1-\pi)^{1-y_i} + \sum_{d=1}^{D}\left(\ln \lambda_{0d} e^{-\lambda_{0d}} + \ln \lambda_{1d} e^{-\lambda_{1d}} + \sum_{i=1}^{n} \ln \frac{e^{-\lambda_{y_i,d}} (\lambda_{y_i,d})^{x_{i,d}}}{x_{i,d}!}\right)$$

$$\nabla_\pi L = \sum_{i=1}^{n}\left(\frac{y_i}{\pi} - \frac{(1-y_i)}{1-\pi}\right) = 0 = \sum_{i=1}^{n}\left(\frac{y_i(1-\pi) - \pi(1-y_i)}{\pi(1-\pi)}\right)$$

$$= \sum_{i=1}^{n} \frac{y_i - \pi}{\pi(1-\pi)} = \sum_{i=1}^{n} \frac{y_i}{\pi(1-\pi)} - \frac{n\pi}{1-\pi} = 0 \qquad \hat{\pi} = \frac{\sum_{i=1}^{n} y_i}{n}$$

(b)
$$\nabla_{\lambda_{y,d}} L = \begin{cases} \text{for } \lambda_{0,d} & -1 + \frac{1}{\lambda_{y,d}} + \sum_{i=1}^{n}(1-y_i)\left(-1 + \frac{x_{i,d}}{\lambda_{y,d}}\right) \\ \text{for } \lambda_{1,d} & -1 + \frac{1}{\lambda_{y,d}} + \sum_{i=1}^{n} y_i\left(-1 + \frac{x_{i,d}}{\lambda_{y,d}}\right) \end{cases} = 0$$

$$\hat{\lambda}_{0,d} = \frac{1 + \sum_{i=1}^{n}(1-y_i)x_{i,d}}{1 + \sum_{i=1}^{n}(1-y_i)} \quad (d = 1, \ldots, D)$$

$$\hat{\lambda}_{1,d} = \frac{1 + \sum_{i=1}^{n} y_i x_{i,d}}{1 + \sum_{i=1}^{n} y_i} \quad (d = 1, \ldots, D)$$

Problem 2:

(a)

y \ y'	0	1
0	2156	631
1	100	1713

$$\text{precision} = \frac{2156 + 1713}{4600} \approx 0.841$$

(b) We can see from the plot that "$\lambda_{y,b}$" for "free" and "!" in "spam" are very different from those in "email". This may indicate that "free" and "!" appear more often in "spam" than they appear in "emails". They can serve as good features in learning process.

(c) k-nn classifier have a higher precision in general in this case than the Naive Bayes Classifier. We may just choose k-nn classifier in this case.