Xiaoqi Zhou (xqzhou@stanford.edu)
06237147

1. **Solution:**
   $\hat{\theta} = \underset{\theta}{argmax} LL(\theta)$
   $LL(\theta) = \sum_{i=1}^{n} \log \lambda e^{-\lambda X_i} = \sum_{i=1}^{n} \log \lambda - \lambda X_i$
   $\dfrac{\delta LL(\lambda)}{\delta \lambda} = \dfrac{n}{\lambda} - \displaystyle\sum_{i=1}^{n} X_i = 0$
   Answer:
   $\hat{\lambda} = \dfrac{n}{\sum_{i=1}^{n} X_i}$

2. **Solution:**
   $Y_i \sim N(\theta_1 X_i + \theta_2, \sigma^2)$
   $LL(\theta_1, \theta_2) = \sum_{i=1}^{n} \log(\frac{1}{\sqrt{2\pi}\sigma}) - \frac{(Y_i - \theta_1 X_i - \theta_2)^2}{2\sigma^2}$
   Answer:
   $LL(\theta_1, \theta_2) = -\frac{n}{2}\log(2\pi) - n\log(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \theta_1 X_i - \theta_2)^2$

3. **Answer:**
   One important assumption of the Naive Bayesian Classifier is all the inputs X_i from the training set it independent. If the first k input of the training are always identical copies.

   $P(X_1, X_2, ..., X_n|Y) \neq \prod_{i=1}^{n} P(X_i|Y)$
   And we have no data can be used as reference to estimate the $P(X_1, X_2, ..., X_n|Y)$, when $X_1, X_2, ...X_k$ are NOT copies of each other.

4. (a) $\dfrac{\delta LL(\theta)}{\theta_1} = \dfrac{\delta LL(\hat{y})}{\hat{y}} \dfrac{\delta \hat{y}}{\delta(\theta_3 g + \theta_4 h)} \dfrac{\delta(\theta_3 g + \theta_4 h)}{\delta g} \dfrac{\delta g}{\delta \theta_1 x} \dfrac{\delta \theta_1 x}{\delta \theta_1}$
   $\dfrac{\delta LL(\theta)}{\theta_1} = \left[\dfrac{y}{\hat{y}} - \dfrac{1-y}{1-\hat{y}}\right] \hat{y}(1-\hat{y})\theta_3 g(1-g)x$
   $\dfrac{\delta LL(\theta)}{\theta_2} = \left[\dfrac{y}{\hat{y}} - \dfrac{1-y}{1-\hat{y}}\right] \hat{y}(1-\hat{y})\theta_4 h(1-h)x$
   $\dfrac{\delta LL(\theta)}{\theta_3} = \left[\dfrac{y}{\hat{y}} - \dfrac{1-y}{1-\hat{y}}\right] \hat{y}(1-\hat{y})g$
   $\dfrac{\delta LL(\theta)}{\theta_4} = \left[\dfrac{y}{\hat{y}} - \dfrac{1-y}{1-\hat{y}}\right] \hat{y}(1-\hat{y})h$