Reading: Sec 7.3, Sec 8.3

1. (MLE for Linear Regression)

   (a) Consider the model $y_i - \boldsymbol{\beta}'\mathbf{z}_i = \varepsilon_i \sim N\left(0, \sigma^2\right),$ where $\mathbf{z}_i = (1, \mathbf{x}_i')'$. Write down the following:
   - The likelihood function
   - The likelihood equations
   - Maximum likelihood estimates

2. (MLE for Logistic Regression)

   (a) Consider the model
   $$\mathbb{P}\left(y_i = 0\right) = p_0\left(\boldsymbol{\beta}, \mathbf{z}_i\right) = \frac{1}{1 + \exp\left(\boldsymbol{\beta}'\mathbf{z}_i\right)}$$
   and
   $$\mathbb{P}\left(y_i = 1\right) = p_1\left(\boldsymbol{\beta}, \mathbf{z}_i\right) = 1 - p_0\left(\boldsymbol{\beta}, \mathbf{z}_i\right)$$
   Write down the following:
   - The likelihood function and the log-likelihood function
   - likelihood equation (denoted by $g\left(\boldsymbol{\beta}\right) = 0$)

   (b) Note that we can not derive the closed form of the MLE here. We need to solve it through numerical methods. One such way is called the Newton-Raphson method. The following problems are somewhat hard, you can skip them if you want.

   i. Find
   $$\frac{\partial^2 L(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}$$

   ii. Let $\mathbf{p}^{(k)} = \left(p_1\left(\boldsymbol{\beta}^{(k)}, \mathbf{z}_i\right), \ldots, p_1\left(\boldsymbol{\beta}^{(k)}, \mathbf{z}_i\right)\right)'$, $W^{(k)} = \operatorname{diag} p_1\left(\boldsymbol{\beta}^{(k)}, \mathbf{z}_i\right)\left(1 - p_1\left(\boldsymbol{\beta}^{(k)}, \mathbf{z}_i\right)\right)$. Use this to simplify
   $$g\left(\boldsymbol{\beta}^{(k)}\right), \qquad \left\{\frac{\partial^2 L(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right\}_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(k)}}$$

   iii. Simplify the following expression using previous results.
   $$\boldsymbol{\beta}^{(k+1)} = \boldsymbol{\beta}^{(k)} - \left\{\frac{\partial^2 L(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right\}^{-1}_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(k)}} \cdot g\left(\boldsymbol{\beta}^{(k)}\right),$$
   What can you find?

   (c) So, actually, steps in (b) are doing:
   $$\tilde{\theta}_0 \longrightarrow \hat{\theta}_0 = \arg\max_{\theta\in\Theta}\left\{\frac{1}{n}\sum_{i=1}^{n}\log f_\theta\left(x_i\right)\right\}$$
   Is the RHS of this equation what we want? What are we missing?