E6690 HW3
Yihang Ding
Yd2459
Yd2459@columbia.edu

P1. $Y = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_1 X_2$

$= -6 + 0.05X_1 + X_2.$

a). $P = \frac{1}{1+e^{-Y}} = \frac{1}{1+e^{0.5}} = 0.378$

b) $P = 0.5 = \frac{1}{1+e^{-(-6+0.05X+3.5)}}$

$X_1 = 50$

P2. $P = P(Y=1 \mid X=4)$

$\cancel{\bar{X} = 0.2\times 0 + 0.8\times 10 = 8.\ s.t.\ X \sim N(8, 6^2)}$

According to Bayes Theorem $\quad X_{issue} \sim N(10, 6^2).\quad X_{non\text{-}issue} \sim N(0, 6^2)$

$$P = P(Y=1 \mid X=4) = \frac{\pi_1 f_1(x)}{\pi_1 f_1(x) + \pi_0 f_0(x)} = \frac{0.8 \cdot \frac{1}{\sqrt{2\pi\cdot 36}} \cdot e^{-\frac{(4-10)^2}{2\cdot 36}}}{0.8 \cdot \frac{1}{\sqrt{2\pi\cdot 36}} \cdot e^{-\frac{(4-10)^2}{2\cdot 36}} + 0.2 \cdot \frac{1}{\sqrt{2\pi\cdot 36}} \cdot e^{-\frac{(4-0)^2}{2\cdot 36}}}$$

$= 0.752.$

P3. $l(\beta_0, \beta_1) = \cancel{\sum \text{sigmoid}(\beta_0+\beta_1 x_i)}$

$$= \sum(-y_i \beta^T \hat{x}_i + \ln(1+e^{\beta^T \hat{x}_i}))$$

$$= \sum_{i=1}^{6}(-y_i(\beta_0+\beta_1 x_i) + \ln(1+e^{\beta_0+\beta_1 x_i}))$$

The goal is to minimize $l(\beta_0, \beta_1)$. i.e. $\beta^* = \arg\min_\beta l(\beta)$

First-Derivative

$$\nabla l(\beta_0, \beta_1) = \begin{bmatrix} \sum_{i=1}^{6}\left(-y_i + \frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}\right) \\ \sum_{i=1}^{6}\left(-x_i y_i + \frac{x_i e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}\right) \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{6} \frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}} - 2 \\ \sum_{i=1}^{6} \frac{x_i e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}} - 1.6 \end{bmatrix}$$

Second-Derivative (Hessian Matrix)

$$J(\beta_0, \beta_1) = \begin{bmatrix} \sum_{i=1}^{6} \frac{e^{\beta_0+\beta_1 x_i}}{(1+e^{\beta_0+\beta_1 x_i})^2}, & \sum_{i=1}^{6} \frac{x_i e^{\beta_0+\beta_1 x_i}}{(1+e^{\beta_0+\beta_1 x_i})^2} \\ \sum_{i=1}^{6} \frac{x_i e^{\beta_0+\beta_1 x_i}}{(1+e^{\beta_0+\beta_1 x_i})^2}, & \sum_{i=1}^{6} \frac{x_i^2 e^{\beta_0+\beta_1 x_i}}{(1+e^{\beta_0+\beta_1 x_i})^2} \end{bmatrix}$$

Step t of Newton Method is $\beta^{t+1} = \beta^t - \left(\frac{\partial^2 l(\beta)}{\partial\beta\partial\beta^T}\right)^{-1} \frac{\partial l(\beta)}{\partial \beta} = \beta^t - J(\beta_0,\beta_1)^{-1}\nabla l(\beta_0,\beta_1)$

in which $\frac{\partial l(\beta)}{\partial \beta} = -\sum_{i=1}^{m} \hat{x}_i (y_i - p_1(x_i;\beta)).$

$$\frac{\partial^2 l(\beta)}{\partial\beta\partial\beta^T} = \sum_{i=1}^{m} \hat{x}_i \hat{x}_i^T p_1(\hat{x}_i;\beta)(1-p_1(\hat{x}_i;\beta))$$

The rest is at the R code part.

P4.

$Cov(Y) = AX(AX)^T$

$= AXX^TA^T$

$= A\Sigma A^T = I$

$\cancel{s.t.\ A\Sigma = A}$.

$\cancel{(\Sigma - I)A = 0}$.

$$\begin{bmatrix} \sigma_1^2-1 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2-1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$\cancel{a_2 = 1.\ a_1 = \frac{\rho\sigma_1\sigma_2}{1-\sigma_1^2}}$

$\cancel{A = \left[\frac{\rho\sigma_1\sigma_2}{1-\sigma \ldots} \right.}$

Do eigen decomposition to Σ, $\Sigma = Q\Lambda Q^T$

$A\Sigma A^T = AQ\Lambda Q^TA^T$

$= AQ\Lambda^{\frac{1}{2}}(Q\Lambda^{\frac{1}{2}})^TA^T = I$.

Thus if $A = (Q\Lambda^{\frac{1}{2}})^{-1}$, $A\Sigma A^T = II^T = I$

$A = (Q\Lambda^{\frac{1}{2}})^{-1} = \Lambda^{-\frac{1}{2}}Q^T$

Decomposition of Σ:

$(\Sigma - \lambda I) = (\sigma_1^2-\lambda)(\sigma_2^2-\lambda) - \rho^2\sigma_1^2\sigma_2^2 = 0$.

We can get λ_1, λ_2 respectively.

P5. As $\hat{\sigma}^2 = \sum_{k=1}^{K} q_k \hat{\sigma}_k^2$

$Var(\hat{\sigma}^2) = \sum_{k=1}^{K} q_k^2 \, Var(\hat{\sigma}_k^2)$, $E(\hat{\sigma}^2) = E(\sum_{k=1}^{K} q_k \hat{\sigma}_k^2) = \sigma^2$

As $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n_k-1}$,

$Var(\frac{(n_k-1)S^2}{\sigma^2}) = Var\, \chi^2_{n_k-1}$

$\frac{(n_k-1)^2}{\sigma^4} Var(S^2) = 2(n_k-1)$

$var(S^2) = \frac{2\sigma^4}{n_k-1}$.

Thus $Var(\hat{\sigma}^2) = \sum_{k=1}^{K} q_k^2 \frac{2\sigma^4}{n_k-1}$

To minimize $Var(\hat{\sigma}^2) = \sum_{k=1}^{K} q_k^2 \frac{2\sigma^4}{n_k-1}$ s.t. $\sum_{k=1}^{K} q_k = 1$.

Use Lagrange Multiplier, get: minimize $L = \sum_{k=1}^{K} q_k^2 \frac{2\sigma^4}{n_k-1} + \lambda(1 - \sum_{k=1}^{K} q_k)$.

$\frac{\partial L}{\partial q} = \sum_{k=1}^{K} \frac{4\sigma^4}{n_k-1} q_k - \sum_{k=1}^{K} \lambda = 0$.

$q_k = \frac{\lambda}{4\sigma^4} (n_k - 1) \propto n_k - 1$.

So $q_k = \frac{n_k-1}{n-K}$

P6. 1. Majority Approach, as $P(p_i > 0.5) = 0.6$, result is Red.

2. Avg Probability, as $Avg(p_i) = 0.45$, result is Green.