Maximum Likelihood Estimator Derivation Logistic Regression

This is the procedure to compute the **Estimates** for the model **Coefficients** .
Continuing from Sigmoid Function Derivation Section we got :

$$P(X_i) = \frac{1}{1+e^S}$$

Let $S = -\vec{X}\vec{\beta}$

Finding the **Likelihood function** for the coefficients β

$$L(\vec{\beta}) = \Pr(Y_1, Y_2 \ldots, Y_n)$$

$$\begin{aligned}
L(\vec{\beta}) &= \prod_{i=1}^{n} \Pr(Y_i = 1) \\
L(\vec{\beta}) &= \prod_{i=1}^{n} P(\vec{X}_i)^{Y_i}(1 - P(\vec{X}_i))^{1-Y_i} \\
\log(L(\vec{\beta})) &= \log\left(\prod_{i=1}^{n} P(\vec{X}_i)^{Y_i}(1 - P(\vec{X}_i))^{1-Y_i}\right) \\
\log(L(\vec{\beta})) &= \sum_{i=1}^{n} Y_i \log(P(\vec{X})) + (1-Y_i)\log(1-P(\vec{X})) \\
\log(L(\vec{\beta})) &= \sum_{i=1}^{n} Y_i \log(\frac{1}{1+e^S}) + (1-Y_i)\log(\frac{e^s}{1+e^S}) \\
\log(L(\vec{\beta})) &= \sum_{i=1}^{n} -Y_i \log(1+e^S) + (1-Y_i)(\log(e^S) - \log(1+e^S)) \\
\log(L(\vec{\beta})) &= \sum_{i=1}^{n} -Y_i \log(1+e^S) + \log(e^S) - \log(1+e^S) - Y_i\log(e^S) + Y_i\log(1+e^S) \\
\log(L(\vec{\beta})) &= \sum_{i=1}^{n} \log(e^S) - \log(1+e^S) - Y_i\log(e^S) \\
\log(L(\vec{\beta})) &= \sum_{i=1}^{n} \log(e^S)(1-Y_i) - \log(1+e^S) \\
\log(L(\vec{\beta})) &= \sum_{i=1}^{n} S(1-Y_i) - \log(1+e^S)
\end{aligned}$$

Derivation with respect to β

$$\begin{aligned}
\log(L(\vec{\beta})) &= \sum_{i=1}^{n} (-\vec{X}\vec{\beta})(1-Y_i) - \log(1+e^{-\vec{X}\vec{\beta}}) \\
\frac{d\log(L(\vec{\beta}))}{d\vec{\beta}} &= \sum_{i=1}^{n} -\vec{X} + \vec{X}Y_i - \frac{1}{1+e^{\vec{X}\vec{\beta}}}(-\vec{X}\ e^{-\vec{X}\vec{\beta}}) \\
\frac{d\log(L(\vec{\beta}))}{d\vec{\beta}} &= \sum_{i=1}^{n} -\vec{X}_i(1-Y_i) + \frac{e^{-\vec{X}\vec{\beta}}}{1+e^{\vec{X}\vec{\beta}}}\vec{X}_i \\
\frac{d\log(L(\vec{\beta}))}{d\vec{\beta}} &= \sum_{i=1}^{n} \vec{X}_i[-(1-Y_i) + \frac{e^{-\vec{X}\vec{\beta}}}{1+e^{\vec{X}\vec{\beta}}}]
\end{aligned}$$

Now we solve for β

$$\sum_{i=1}^{n} \vec{X}_i \left[-(1-Y_i) + \frac{e^{-\vec{X}\vec{\beta}}}{1+e^{\vec{X}\vec{\beta}}}\right] = 0$$

Derivation Notes

- We log the likelihood so the products turn into sums which are easier to calculate
- We derive with respect for β so we can find where the likelihood will maximize with the β as a parameter

Conclusion

- Unlike Linear RegressionOrdinary Least Squares, the **Likelihood function** in the logistic regression in **nonlinear** in the parameters β, the gradient of the log- likelihood do not yield a closed-for solution for β
- We use **iterative numerical optimization methods**