Lab Solutions 09

1. I produced the derivation of the logistic score function for you; make sure that you can reproduce these steps yourself. Parts of this may be on the next exam.

2. Following the instructions in the RMarkdown file to implement logistic regression using gradient descent.

See markdown solutions.

3. Find a formula for the second derivatives of the loss function:

$$\frac{\partial^2 l(b)}{\partial b_k \partial b_j} \tag{7.1}$$

Using the chain run, this is mostly a straightforward calculation if you use the compact form of p_i (see supplementary file on class website):

$$\begin{aligned}
\frac{\partial^2 l(y)}{\partial \beta_k \partial \beta_j} &= \sum_i x_{i,j} \frac{\partial p_i}{\partial \beta_k} &(7.2)\\
&= \sum_i x_{i,j} - 1 \cdot \left(\frac{1}{1 + e^{-x^t \beta}} \right)^2 (-x_{i,k}) \cdot e^{-x_i^t \beta} &(7.3)\\
&= \sum_i x_{i,j} x_{i,k} \cdot \left(\frac{e^{-x^t \beta}}{1 + e^{-x^t \beta}} \right) \cdot \left(\frac{1}{1 + e^{-x^t \beta}} \right) &(7.4)\\
&= \sum_i x_{i,j} x_{i,k} p_i (1 - p_i). &(7.5)
\end{aligned}$$

The last line is also explained better in the supplementary file.

4. The Hessian matrix is a square matrix of partial derivatives. That is, $H_{k,j} = \frac{\partial^2 l(b)}{\partial b_k \partial b_j}$. From the result in the previous question, understand that we can write:

$$H = X^t D X \tag{7.6}$$

Where D is a diagonal matrix:

$$D_{i,i} = p_i (1 - p_i). \tag{7.7}$$

If you do the simplification that I did of p_i and $1 - p_i$, this follows directly.

5. The "proper" way to logistic regression is by replacing the learning rate ρ with the Hessian matrix evaluated at the current value of b. That is:

$$b^{(t+1)} = b^{(t)} - H(b^{(t)}) \cdot \nabla_b l(b^{(t)}). \tag{7.8}$$

This corresponds to doing a quadratic approximation at $b^{(t)}$ and moving directly to the minimizing point. Implement this iteration in the RMarkdown file and compare the convergence rate to the gradient descent implementation.

See markdown solutions.