Assignment 1: Written Exercises

Kulin Seth

February 22, 2016

1 Question 1

Variance of a sum. Show that the variance of a sum is

$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2 * \mathrm{Cov}[X, Y]$

where $\mathrm{Cov}[X, Y]$ is the covariance between random variables X and Y.

2 Question 2

3 Question 3

Gradient and Hessian of log-likelihood for logistic regression.

3.1 a

Let

$$\rho(a) = \frac{1}{1 + \exp^{-a}} \tag{1}$$

be the sigmoid function. Show that :

$$\frac{\mathrm{d}}{\mathrm{d}a}\rho(a) = \rho(a)(1\rho(a)) \tag{2}$$

Proof
Differentiating $\rho(a)$ w.r.t a we get :
L.H.S

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}a}\rho(a) &= \frac{\mathrm{d}}{\mathrm{d}a}\left(\frac{1}{1+e^{-a}}\right) \\
&= \left(\frac{-1}{(1+e^{-a})^2}\right)\frac{\mathrm{d}}{\mathrm{d}a}\left(1+e^{-a}\right) && \text{Reciprocal rule:}\frac{1}{f} \to \frac{-f\prime}{f^2} \\
&= \left(\frac{-1}{(1+e^{-a})^2}\right)\left(-e^{-a}\right) && \text{Derivative of:}e^{-a} \to -e^{-a} \\
&= \frac{e^{-a}}{(1+e^{-a})^2} && \text{Simplifying}
\end{aligned}$$

Substituting $\rho(a)$ on the R.H.S we get:
R.H.S

$$\begin{aligned}
\rho(a)(1-\rho(a)) &= \frac{1}{1+e^{-a}}\left(1 - \frac{1}{1+e^{-a}}\right) && \text{(substituting)} \\
&= \frac{\exp^{-a}}{(1+\exp^{-a})^2} && \text{(simplifying)}
\end{aligned}$$

$R.H.S \iff L.H.S$ hence proved.

3.2 b

Using the previous result and the chain rule of calculus, derive the expression for the gradient of the log likelihood given in HTF Eqn. 4.21.

Equation (4.20) states:

$$l(\beta) = \sum_{i=1}^{N} \{y_i \log p(x_i;\beta) + (1-y_i)\log(1-p(x_i;\beta))\} \tag{3}$$

$$l(\beta) = \sum_{i=1}^{N} \left\{ y_i \beta^T x_i - log(1+\exp^{\beta^T x_i}) \right\} \tag{4}$$

We can get from (**??**) to (**??**) by substituting

$$p(x_i;\beta) = \frac{e^{\beta^T x_i}}{1+e^{\beta^T x_i}}$$

PROOF

To Prove:

$$\frac{\partial}{\partial\beta}(l(\beta)) = \sum_{i=1}^{N} x_i(y_i - p(x_i;\beta))$$

Differentiating (**??**) w.r.t to β on both sides we get:

$$\begin{aligned}
\frac{\partial}{\partial\beta}(l(\beta)) &= \frac{\partial}{\partial\beta}(\sum_{i=1}^{N}(y_i\beta^T x_i - log(1+\exp^{\beta^T x_i}))) \\
&= \sum_{i=1}^{N}(\frac{\partial}{\partial\beta}(y_i\beta^T x_i)) - (\frac{\partial}{\partial\beta}(log(1+\exp^{\beta^T x_i}))) \qquad \text{moving derivative inside} \\
&= \sum_{i=1}^{N} y_i x_i - \frac{1}{1+e^{\beta^T x_i}}(x_i e^{\beta^T x_i})
\end{aligned}$$

3.3 c

As noted in HTF Eqn. 4.25, the Hessian matrix for the log likelihood can be written (up to a sign) as $X^T W X$. Prove that this matrix is positive definite.