Computer Practical 2

Statistics 2

Dom Hutchinson

Consider the random variables $Y_i \overset{\text{ind}}{\sim} \text{Bernoulli}(\sigma(\theta^T x_i))$ for $i \in [1, n]$ where x_i are observed d-dimensional real vectors of explanatory variables and σ is the standard logistic function.

Question 1

By definition we have $L(\theta; \mathbf{y}) \propto \prod_{i=1}^n f_{Y_i}(y_i; \theta)$.

Thus, in this case $L(\theta; \mathbf{y}) \propto \prod_{i=1}^n \sigma(\theta^T x_i)^{y_i}(1 - \sigma(\theta^T x_i))^{1-y_i}$ We have

$$\begin{array}{rcl} \ell(\theta; \mathbf{y}) & := & \ln L(\theta; \mathbf{y}) \\ & = & \ln \left[L(\theta; \mathbf{y}) \propto \prod_{i=1}^n \sigma(\theta^T x_i)^{y_i}(1 - \sigma(\theta^T x_i))^{1-y_i} \right] \\ & = & \sum_{i=1}^n \ln \sigma(\theta^T x_i)^{y_i} + \ln[1 - \sigma(\theta^T x_i)]^{1-y_i} \\ & = & \sum_{i=1}^n y_i \ln \sigma(\theta^T x_i) + (1 - y_i) \ln[1 - \sigma(\theta^T x_i)] \end{array}$$

Question 2

Consider $\frac{\partial}{\partial \theta_j} \sigma(\theta^T x_i)$ we have

$$\begin{array}{rcl} \frac{\partial}{\partial \theta_j} & = & \left[\frac{\partial}{\partial \theta_j} \sigma(\theta^T x_i) \right] \cdot \left[\frac{\partial}{\partial \theta_j} \theta^T x_i \right] \text{ by chain rule} \\ & = & \sigma(\theta^T x_i)(1 - \sigma(\theta^T x_i)) x_{ij} \end{array}$$

Note that $\frac{\partial}{\partial \theta_j} \theta_T = (\delta_{ij} : i \in [1, n])$, the vectors of all zeros except at position j where it is 1.
Thus $\frac{\partial}{\partial \theta_j} \theta_T x_i = x_{ij}$.
Now consider $\frac{\partial}{\partial \theta_j} \ell(\theta; \mathbf{y})$. Then

$$\begin{array}{rcl} \frac{\partial}{\partial \theta_j} \ell(\theta; \mathbf{y}) & = & \sum_{i=1}^n y_i \frac{\partial}{\partial \theta_j} \ln \sigma(\theta^T x_i) + (1 - y_i) \frac{\partial}{\partial \theta_j} \ln[1 - \sigma(\theta^T x_i)] \\ & = & \sum_{i=1}^n y_i \frac{x_{ij} \sigma(\theta^T x_i)[1 - \sigma(\theta^T x_i)]}{\sigma(\theta^T x_i)} + (1 - y_i) \frac{-x_{ij} \sigma(\theta^T x_i)[1 - \sigma(\theta^T x_i)]}{1 - \sigma(\theta^T x_i)} \\ & = & \sum_{i=1}^n y_i x_{ij}[1 - \sigma(\theta^T x_i)] - (1 - y_i) x_{ij} \sigma(\theta^T x_i) \\ & = & \sum_{i=1}^n y_i x_{ij} - y_i x_{ij} \sigma(\theta^T x_i) - x_{ij} \sigma(\theta^T x_i) + y_i x_{ij} \sigma(\theta^T x_i) \\ & = & \sum_{i=1}^n [y_i - \sigma(\theta^T x_i)] x_{ij} \end{array}$$

Question 3

Consider $\frac{\partial^2}{\partial\theta_j\partial\theta_k}\ell(\theta;\mathbf{y})$. Then

$$\begin{aligned}
\frac{\partial^2}{\partial\theta_j\partial\theta_k}\ell(\theta;\mathbf{y}) &= \frac{\partial}{\partial\theta_k}\left[\frac{\partial}{\partial\theta_j}\ell(\theta;\mathbf{y})\right] \\
&= \frac{\partial}{\partial\theta_k}\left[\sum_{i=1}^{n}[y_i - \sigma(\theta^T x_i)]x_{ij}\right] \\
&= \sum_{i=1}^{n} 0 - x_{ij}\frac{\partial}{\partial\theta_k}\sigma(\theta^T x_i) \\
&= -\sum_{i=1}^{n} x_{ij}\sigma(\theta^T x_i)[1-\sigma(\theta^T x_i)]x_{ik} \text{ by result in Question 2} \\
&= -\sum_{i=1}^{n} \sigma(\theta^T x_i)[1-\sigma(\theta^T x_i)]x_{ij}x_{ik}
\end{aligned}$$