Exercise Sheet 6

Robin Schmidt
Probabilistic Inference & Learning

November 21, 2018

Generalized Linear Models

1. Newton Optimization

(a)

Proof. First we recall the taylor expansion in vector notation:

$$f(x+\delta x) = f(x) + \sum_{j=1}^{N} \frac{\partial f(x)}{\partial x_j}\delta x_j + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\delta x_i \delta x_j + ...$$
$$f(x+\delta x) = f(x) + \delta^T x \nabla f(x) + \frac{1}{2}\delta^T x H \delta x + ...$$

Using the notation stated in the exercise and constructing a quadratic approximation we get:

$$L(f_0+\delta) = \widetilde{L}(f_0+\delta) = L(f_0) + \delta^T \nabla L(f_0) + \frac{1}{2}\delta^T B(f_0)\delta$$
$$f_1 = f_0 + \delta \quad \text{which leads to:} \quad \delta = f_1 - f_0$$

Taking the derivative in respect to δ equal to zero and substituting $\delta = f_1 - f_0$ gives the following:

$$\begin{aligned} 0 \stackrel{!}{=} \frac{d}{d\delta}\widetilde{L}(f_0+\delta) &= \nabla L(f_0) + B(f_0)\delta \\ -B(f_0)(f_1 - f_0) &= \nabla L(f_0) \\ B(f_0)(f_0 - f_1) &= \nabla L(f_0) \\ f_0 - f_1 &= B^{-1}(f_0)\nabla L(f_0) \\ f_1 &= f_0 - B^{-1}(f_0)\nabla L(f_0) \end{aligned}$$

□

Which shows that the minimum of this approximation lies at $f_1 = f_0 - B^{-1}(f_0)\nabla L(f_0)$ which was to be shown.

(b)

Rewriting the second-order expansion from above and setting the gradient to zero gives the following:

$$
\begin{aligned}
L(w+\epsilon) = \widetilde{L}(w+\epsilon) &= L(\phi^T w) + \epsilon^T \nabla L(\phi^T w) + \frac{1}{2}\epsilon^T B(\phi^T w)\epsilon \\
0 &\overset{!}{=} \frac{d}{d\epsilon}\widetilde{L}(w+\epsilon) = \nabla L(\phi^T w) + B(\phi^T w)\epsilon \\
-B(\phi^T w)(w - w_0) &= \nabla L(\phi^T w) \\
B(\phi^T w)(w_0 - w) &= \nabla L(\phi^T w) \\
w_0 - w &= B^{-1}(\phi^T w)\nabla L(\phi^T w) \\
w &= w_0 - B^{-1}(\phi^T w)\nabla L(\phi^T w)
\end{aligned}
$$

This leads to the Newton step in w with $w = w_0 - B^{-1}(\phi^T w)\nabla L(\phi^T w)$, which was asked in the exercise.

2. Logistic Link Function

$$\frac{\partial \ log\ \sigma(yf)}{\partial f} = \frac{ye^{yf}}{(e^{yf}+1)^2}$$