Solution to small exercise (Lecture 2, Day 2)

Helene Charlotte Wiese Rytgaard

September 24, 2021

Define:

(i) Log-likelihood loss function:

$$\mathscr{L}(f)(O) = -\big(Y \log(f(A,X)) + (1-Y)\log(1-f(A,X))\big).$$

(ii) Logistic regression model:

$$f_\varepsilon(A,X) = \operatorname{expit}\big(\operatorname{logit}(f(A,X)) + \varepsilon H(A,X)\big),$$

with the so-called "clever covariate",

$$H(A,X) := \frac{2A-1}{\pi(A \mid X)},$$

We verify that (i)–(ii) fulfill

$$(1) \quad f_{\varepsilon=0} = f \qquad (2) \quad \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} \mathscr{L}(f_\varepsilon)(O) = \phi^*(f,\pi)(O).$$

(1) follows immediately. For (2), first recall that

$$\operatorname{expit}(x) = \frac{e^x}{1+e^x} = \frac{1}{1+e^{-x}}, \qquad \operatorname{logit}(x) = \log\left(\frac{x}{1-x}\right),$$

such that

$$\operatorname{expit}(\operatorname{logit}(x)) = x,$$

and

$$\begin{aligned} \operatorname{expit}(-\operatorname{logit}(x)) &= \operatorname{expit}(-\log(x) + \log(1-x)) \\ &= \operatorname{expit}(\log(1-x) - \log(x)) \\ &= \operatorname{expit}(\operatorname{logit}(1-x)) = 1-x. \end{aligned}$$

Furthermore, it is easily seen that

$$\begin{aligned} \log(\operatorname{expit}(x)) &= -\log(1+e^{-x}), \\ \log(1-\operatorname{expit}(x)) &= -\log(1+e^{x}), \end{aligned}$$

so that

$$\begin{aligned}
\frac{d}{dx}\log(\operatorname{expit}(x)) &= \frac{e^{-x}}{1+e^{-x}} = \operatorname{expit}(-x), \\
\frac{d}{dx}\log(1-\operatorname{expit}(x)) &= -\frac{e^{x}}{1+e^{x}} = -\operatorname{expit}(x).
\end{aligned}$$

Now we can differentiate the composite functions

$$\begin{aligned}
\frac{d}{d\varepsilon}\log(\operatorname{expit}(\operatorname{logit}(x)+\varepsilon h)) &= h\operatorname{expit}(-(\operatorname{logit}(x)+\varepsilon h)), \\
\frac{d}{d\varepsilon}\log(1-\operatorname{expit}(\operatorname{logit}(x)+\varepsilon h)) &= -h\operatorname{expit}(\operatorname{logit}(x)+\varepsilon h),
\end{aligned}$$

where setting $\varepsilon = 0$ gives

$$\begin{aligned}
\frac{d}{d\varepsilon}\log(\operatorname{expit}(\operatorname{logit}(x)+\varepsilon h))\Big|_{\varepsilon=0} &= h\operatorname{expit}(-(\operatorname{logit}(x))) = h(1-x), \\
\frac{d}{d\varepsilon}\log(1-\operatorname{expit}(\operatorname{logit}(x)+\varepsilon h))\Big|_{\varepsilon=0} &= -h\operatorname{expit}(\operatorname{logit}(x)) = hx.
\end{aligned}$$

Applying these steps to $\mathscr{L}(f_\varepsilon)$ now gives:

$$\begin{aligned}
\frac{d}{d\varepsilon}\Big|_{\varepsilon=0}\mathscr{L}(f_\varepsilon)(O) &= YH(A,X)(1-f(A,X)) - (1-Y)H(A,X)f(A,X) \\
&= YH(A,X) - H(A,X)f(A,X) = H(A,X)\big(Y-f(A,X)\big).
\end{aligned}$$