Situation: We have n data points $(x_i, y_i),\ i = 1, ..., n$ with $x_i \in \mathbb{R}^d$ and $y \in \{-1, 1\}$.

Probit Model: y_i is a realization of the random variable Y_i. $Y_1, ..., Y_n$ are independent. The distribution of Y_i is as follows:

$$P(Y_i = 1|x_i;\beta) = \Phi(x_i^T\beta)$$
$$P(Y_i = -1|x_i;\beta) = 1 - \Phi(x_i^T\beta)$$

where $\beta \in \mathbb{R}^d$. It follows that

$$P(Y_i = y_i|x_i;\beta) = \Phi(y_i x_i^T\beta)$$

Likelihood: The likelihood of a parameter vector β is given as follows:

$$L(\beta) = \prod_{i=1}^{n} P(Y_i = y_i|x_i;\beta) = \prod_{i=1}^{n} \Phi(y_i x_i^T\beta)$$

The negative log-likelihood that we wish to minimize is:

$$\mathcal{L}(\beta) = -\sum_{i=1}^{n} \log \Phi(y_i x_i^T\beta)$$

The weighted case: We introduce sample weights $w_i \in \mathbb{R}_{>0}$ comprising a weight vector $w \in \mathbb{R}^n_{>0}$. Further, let $g(z) = -\log \Phi(-z)$. The objective function now becomes:

$$f_w(\beta) = \sum_{i=1}^{n} w_i g(-y_i x_i^T\beta)$$

To make the notation easier, we define $z_i = -y_i x_i^T$ and introduce the matrix $Z \in \mathbb{R}^{n\times d}$ with row vectors $Z_i = z_i$. This gives us:

$$f_w(\beta) = \sum_{i=1}^{n} w_i g(z_i\beta)$$

Lemma 1. *Let $g(z) = -\log \Phi(-z)$. Then it holds for all $z \geq 0$ that:*

$$\frac{1}{2}z^2 \leq g(z)$$

For all $z \geq 2$ it holds that:

$$g(z) \leq 2z^2$$

Proof. TODO. □

Definition 1. *Let $Z \in \mathbb{R}^{n\times d}$. Then we define*

$$\mu_w(Z) = \sup_{\beta\in\mathbb{R}^d\setminus\{0\}} \frac{\left\|(\sqrt{D_w}Z\beta)^+\right\|_2^2}{\left\|(\sqrt{D_w}Z\beta)^-\right\|_2^2} = \sup_{\beta\in\mathbb{R}^d\setminus\{0\}} \frac{\left\|(\sqrt{D_w}Z\beta)^-\right\|_2^2}{\left\|(\sqrt{D_w}Z\beta)^+\right\|_2^2}$$

Z weighted by w is called μ-complex if $\mu_w(Z) \le \mu$.

Lemma 2. *Let $Z \in \mathbb{R}^{n\times d}$ weighted by $w \in \mathbb{R}^n_{>0}$ be μ-complex. Let U be an orthonormal basis for the columnspace of $\sqrt{D_w}Z$. If for index i, the supreme β in (TODO) satisfies $2 \le z_i\beta$, then $w_i g(z_i\beta) \le 4\|U_i\|_2^2(1+\mu)f_w(\beta)$.*

Proof. Let $\sqrt{D_w}Z = UR$, where U is an orthonormal basis for the columnspace of $\sqrt{D_w}Z$. It follows from $2 \le z_i\beta$ and from the monotonicity of g that

$$\begin{aligned}
w_i g(z_i\beta) &= w_i g\left(\frac{\sqrt{w_i}z_i\beta}{\sqrt{w_i}}\right) = w_i g\left(\frac{U_iR\beta}{\sqrt{w_i}}\right) \le w_i g\left(\frac{\|U_i\|_2\|R\beta\|_2}{\sqrt{w_i}}\right)\\
&= w_i g\left(\frac{\|U_i\|_2\|UR\beta\|_2}{\sqrt{w_i}}\right) = w_i g\left(\frac{\|U_i\|_2\|\sqrt{D_w}Z\beta\|_2}{\sqrt{w_i}}\right)\\
&\le 2\|U_i\|_2^2\|\sqrt{D_w}Z\beta\|_2^2 \le 2\|U_i\|_2^2(1+\mu)\|(\sqrt{D_w}Z\beta)^+\|_2^2\\
&= 2\|U_i\|_2^2(1+\mu)\sum_{j:\ \sqrt{w_j}z_j\beta\ge 0} w_j(z_j\beta)^2\\
&\le 4\|U_i\|_2^2(1+\mu)\sum_{j:\ \sqrt{w_j}z_j\beta\ge 0} w_j g(z_j\beta)\\
&\le 4\|U_i\|_2^2(1+\mu)\sum_{j=1}^{n} w_j g(z_j\beta)\\
&= 4\|U_i\|_2^2(1+\mu)f_w(\beta)
\end{aligned}$$

□

Lemma 3. *Let $Z \in \mathbb{R}^{n\times d}$ weighted by $w \in \mathbb{R}^n_{>0}$ be μ-complex. If for index i, the supreme β in (TODO) satisfies $z_i\beta \le 2$, then $w_i g(z_i\beta) \le \frac{w_i}{\mathcal{W}}(80+16\mu)f_w(\beta)$.*

Proof. Let $K^- = \{j \in [n] \mid z_j\beta \le -1\}$ and $K^+ = \{j \in [n] \mid z_j\beta > -1\}$. Note that $g(-1) > \frac{1}{10}$ and $g(z_i\beta) \le g(2) < 4$. Also, $\sum_{j\in K^+} w_j + \sum_{j\in K^-} w_j = \mathcal{W}$.
Thus, if $\sum_{j\in K^+} w_j \ge \frac{1}{2}\mathcal{W}$ then

$$f_w(\beta) = \sum_{j=1}^{n} w_j g(z_j\beta) \ge \sum_{j\in K^+} w_j g(z_j\beta) \ge \frac{\sum_{j\in K^+} w_j}{10} \ge \frac{\mathcal{W}}{20} = \frac{\mathcal{W}}{20w_i}w_i \ge \frac{\mathcal{W}}{80w_i}w_i g(z_i\beta)$$

If on the other hand $\sum_{j\in K^+} w_j < \frac{1}{2}\mathcal{W}$, then $\sum_{j\in K^-} w_j \geq \frac{1}{2}\mathcal{W}$. Thus

$$\begin{aligned}
f_w(\beta) &= \sum_{j=1}^{n} w_j g(z_j\beta) \geq \sum_{j:\ z_j\beta>0} w_j g(z_j\beta) \geq \frac{1}{2}\sum_{j:\ z_j\beta>0} w_j (z_j\beta)^2 \\
&= \frac{1}{2}\|(\sqrt{D_w}Z\beta)^+\|_2^2 \geq \frac{1}{2\mu}\|(\sqrt{D_w}Z\beta)^-\|_2^2 \\
&= \frac{1}{2\mu}\sum_{j:\ z_j\beta<0} w_j (z_j\beta)^2 \\
&\geq \frac{1}{2\mu}\sum_{j\in K^-} w_j (z_j\beta)^2 \\
&\geq \frac{1}{2\mu}\sum_{j\in K^-} w_j \\
&\geq \frac{\mathcal{W}}{4\mu} \\
&\geq \frac{\mathcal{W}}{16\mu w_i} w_i g(z_i\beta)
\end{aligned}$$

Adding both bounds, we get that for $z_i\beta \leq 2$:

$$w_i g(z_i\beta) \leq f_w(\beta)\frac{80 w_i}{\mathcal{W}} + f_w(\beta)\frac{16\mu w_i}{\mathcal{W}} = \frac{w_i}{\mathcal{W}}(80+16\mu) f_w(\beta)$$

□

Lemma 4. *Let $Z \in \mathbb{R}^{n\times d}$ weighted by $w \in \mathbb{R}^n_{>0}$ be μ-complex. Let U be an orthonormal basis for the columnspace of $\sqrt{D_w}Z$. For each $i \in [n]$, the sensitivity of $g_i(\beta) = g(z_i\beta)$ is bounded by $\varsigma_i \leq s_i = (80+16\mu)(\|U_i\|_2^2 + \frac{w_i}{\mathcal{W}})$. The total sensitivity is bounded by $\mathfrak{S} \leq$ something.*

Proof.

$$\begin{aligned}
\varsigma_i &= \sup_\beta \frac{w_i g(z_i\beta)}{f_w(\beta)} \leq \sup_\beta \frac{4\|U_i\|_2^2(1+\mu) f_w(\beta) + \frac{w_i}{\mathcal{W}}(80+16\mu) f_w(\beta)}{f_w(\beta)} \\
&= 4\|U_i\|_2^2(1+\mu) + \frac{w_i}{\mathcal{W}}(80+16\mu) \\
&\leq \|U_i\|_2^2(80+16\mu) + \frac{w_i}{\mathcal{W}}(80+16\mu) \\
&= (80+16\mu)(\|U_i\|_2^2 + \frac{w_i}{\mathcal{W}})
\end{aligned}$$

$$\begin{aligned}
\mathfrak{S} = \sum_{i=1}^{n} \varsigma_i &\leq (80+16\mu) \sum_{i=1}^{n} \|U_i\|_2^2 + \frac{w_i}{\mathcal{W}} \\
&= (80+16\mu)(\|U\|_F + 1) \\
&\leq (80+16\mu)(\sqrt{d}+1) \\
&\leq 96\mu(\sqrt{d}+1) \\
&\leq 192\mu\sqrt{d}
\end{aligned}$$

□

References