EECS 553 Homework 8 Solution (FA24)

1. Logistic Regression with Label Proportion Shift

a. Grading Rubrics:

- Full points for equating $\Phi_k(x)$ and $\Phi'_k(x)$ to derive the correct final expression of $p'(y = k|x)$.
- Minus 0.5 pts for each minor mistake, e.g., sign error, missing symbol/variable, etc.
- 0 pts if no effort or entirely wrong.

Start by using Bayes' Rule for pre- and post-shift class-conditional likelihoods:

$$\phi_k(x) = \frac{p(y = k|x)p(x)}{\pi_k} \tag{1}$$

$$\phi'_k(x) = \frac{p'(y = k|x)p'(x)}{\pi'_k} \tag{2}$$

Since $\phi_k = \phi'_k$, we can equate the above expressions and solve for $p'(y = k|x)$ as follows:

$$p'(y = k|x) = p(y = k|x)\frac{p(x)\pi'_k}{p'(x)\pi_k} \tag{3}$$

We can then use the given identity $p'(x)/p(x) = \frac{\pi'_0}{\pi_0}p(y = 0|k) + \frac{\pi'_1}{\pi_1}p(y = 1|x)$. Substituting into 3 we have the final result:

$$p'(y = k|x) = \frac{\frac{\pi'_k}{\pi_k}p(y = k|x)}{\frac{\pi'_0}{\pi_0}p(y = 0|x) + \frac{\pi'_1}{\pi_1}p(y = 1|x)}, \quad k = 0, 1 \tag{4}$$

b. Grading Rubrics:

- 1.5 pts each for correct final expressions of the incomplete ($\ell(\pi') = \log p(\mathbb{X}|\pi')$) and complete ($\log p(\mathbb{X}, \mathbb{Z}|\pi')$) data log-likelihood functions.
- Minus 0.5 pts for each minor mistake, e.g., sign error, missing symbol/variable, missing summation, etc.
- 0 pts if no effort or entirely wrong.

Starting with the incomplete data log likelihood $\ell(\pi') = \log p(\mathbb{X}|\pi')$.
By the independence of the data samples:

$$\ell(\pi') = \log \prod_{i=1}^{m} p(x_i|\pi')$$

Next using the law of total probability:

$$\ell(\pi') = \log \prod_{i=1}^{m} \phi'_0(x_i)\pi'_0 + \phi'_1(x_i)\pi'_1$$

Finally, distribute the log for the simplified incomplete data log likelihood:

$$\ell(\pi') = \sum_{i=1}^{m} \log \left[\phi_0'(x_i)\pi_0' + \phi_1'(x_i)\pi_1'\right]$$

Next, we'll find the complete data log likelihood, again starting from the independence of data samples:

$$\log p(\mathbb{X}, \mathbb{Z}|\pi') = \log \prod_{i=1}^{m} p(x_i, y_i|\pi')$$

Then use the law of total probability over the labels, and recall that $y_i \in \{0, 1\}$.

$$\log p(\mathbb{X}, \mathbb{Z}|\pi') = \log \prod_{i=1}^{m} (\phi_0'(x_i)\pi_0')^{1-y_i}(\phi_1'(x_i)\pi_1')^{y_i}$$

$$\log p(\mathbb{X}, \mathbb{Z}|\pi') = \sum_{i=1}^{m} (1 - y_i)(\log \pi_0' + \log \phi_0'(x_i)) + y_i(\log \pi_1' + \log \phi_1'(x_i))$$

Giving the simplified complete data log likelihood.

c. Grading Rubrics:

- 1.5 pts each for correct E-step and M-step.
- Minus 0.5 pts for each minor mistake, e.g., sign error, missing symbol/variable, missing summation, etc.
- 0 pts if no effort or entirely wrong.

First, we'll derive the E-step for $Q(\pi', \pi'^{(t)}) = \mathbb{E}\left[\log p(\mathbb{X}, \mathbb{Z}|\pi')|\mathbb{X}, \pi'^{(t)}\right]$.
Using the linearity of expectation and the derived expression in (b) for the complete data log-likelihood:

$$\begin{aligned} Q(\pi', \pi'^{(t)}) = \sum_{i=1}^{m} &(1 - \mathbb{E}[y_i|x_i, \pi'^{(t)}])(\log \pi_0' + \log \phi_0'(x_i)) \\ &+ \mathbb{E}[y_i|x_i, \pi'^{(t)}]((\log \pi_1' + \log \phi_1'(x_i)) \end{aligned}$$

We can find $\mathbb{E}[y_i|x_i, \pi'^{(t)}]$ using the definition of expectation, recalling that $y_i \in \{0, 1\}$.

$$\begin{aligned} \mathbb{E}[y_i|x_i, \pi'^{(t)}] &= 0 \cdot p(y_i = 0|x_i, \pi'^{(t)}) + 1 \cdot p(y_i = 1|x_i, \pi'^{(t)}) \\ &= p(y_i = 1|x_i, \pi'^{(t)}) \end{aligned}$$

Either of the following are acceptable solutions:

$$\begin{aligned} Q(\pi', \pi'^{(t)}) = \sum_{i=1}^{m} &(1 - p(y_i = 1|x_i, \pi'^{(t)}))(\log \pi_0' + \log \phi_0'(x_i)) \\ &+ p(y_i = 1|x_i, \pi'^{(t)})((\log \pi_1' + \log \phi_1'(x_i)) \end{aligned}$$

$$\begin{aligned} Q(\pi', \pi'^{(t)}) = \sum_{i=1}^{m} &p(y_i = 0|x_i, \pi'^{(t)})(\log \pi_0' + \log \phi_0'(x_i)) \\ &+ p(y_i = 1|x_i, \pi'^{(t)})((\log \pi_1' + \log \phi_1'(x_i)) \end{aligned}$$

Onto the M-step. Note that we can substitute $\pi'_0 = 1 - \pi'_1$, and solve the optimization for π'_1.

$$\begin{aligned}\frac{\partial Q(\pi', \pi'^{(t)})}{\partial \pi'_1} &= \sum_{i=1}^{m} \frac{1 - p'^{(t)}(y_i = 1|x_i)}{1 - \pi'_1} - \frac{p^{(t)}(y_i = 1|x_i)}{\pi'_1} \\ &= 0\end{aligned}$$

$$\begin{aligned}\sum_{i=1}^{m} \frac{1}{1 - \pi'_1} &= \sum_{i=1}^{m} \frac{p'^{(t)}(y_i = 1|x_i)}{1 - \pi'_1} - \frac{p^{(t)}(y_i = 1|x_i)}{\pi'_1} \\ &= \sum_{i=1}^{m} \frac{p'^{(t)}(y_i = 1|x_i)(\pi'_1 + 1 - \pi'_1)}{\pi'_1(1 - \pi'_1)}\end{aligned}$$

$$\begin{aligned}m\pi'_1 \frac{1 - \pi'_1}{1 - \pi'_1} &= \sum_{i=1}^{m} p'^{(t)}(y_i = 1|x_i) \\ \pi'_1 &= \frac{1}{m} \sum_{i=1}^{m} p'^{(t)}(y_i = 1|x_i)\end{aligned}$$

We then have the update $\pi'^{(t+1)}_k = \frac{1}{m}\sum_{i=1}^{m} p'^{(t)}(y_i = k|x_i)$; or using shorthand we arrive at the form given in the problem $\pi'^{(t+1)} = \frac{1}{m}\sum_{i=1}^{m} p'^{(t)}(y_i = 1|x_i)$.

d. Grading Rubrics:

- 1 pts for each accuracy (adjusted, unadjusted, clairvoyant LRs) within $\pm 2\%$.
- 0 pts if no effort or entirely wrong.

(i.) Adjusted LR accuracy: 90%

(ii.) Unadjusted LR accuracy: 83%

(iii.) Clairvoyant LR accuracy: 90%

2. Linear Regression with Laplacian Likelihood (5 points)

Grading Rubrics:

- 2 pts for correctly writing the Laplacian likelihood function.
- 2 pts for deriving the negative log-likelihood and identifying the corresponding loss function.
- 1 pt for concluding the equivalence to empirical risk minimization.
- Minus 0.5 pts for each minor mistake, e.g., incorrect constants, missing absolute value, etc.
- 0 pts if no effort or entirely wrong.

Start by expressing the likelihood of each observation under the Laplacian distribution:

$$p(y_i|\boldsymbol{x}_i, \boldsymbol{w}) = \frac{1}{2b} \exp\left(-\frac{|y_i - \boldsymbol{w}^T \boldsymbol{x}_i|}{b}\right)$$

where $b > 0$ is the scale parameter of the Laplacian distribution.

The likelihood for the entire dataset is the product of individual likelihoods:

$$L(\boldsymbol{w}) = \prod_{i=1}^{n} p(y_i|\boldsymbol{x}_i, \boldsymbol{w}) = \left(\frac{1}{2b}\right)^n \exp\left(-\frac{1}{b}\sum_{i=1}^{n}|y_i - \boldsymbol{w}^T\boldsymbol{x}_i|\right)$$

To perform maximum likelihood estimation, we take the logarithm of the likelihood to obtain the log-likelihood:

$$\log L(\boldsymbol{w}) = -n\log(2b) - \frac{1}{b}\sum_{i=1}^{n}|y_i - \boldsymbol{w}^T\boldsymbol{x}_i|$$

Maximizing the log-likelihood is equivalent to minimizing the negative log-likelihood:

$$\text{Negative Log-Likelihood} = n\log(2b) + \frac{1}{b}\sum_{i=1}^{n}|y_i - \boldsymbol{w}^T\boldsymbol{x}_i|$$

Since $n\log(2b)$ is a constant with respect to $\boldsymbol{w}$, the optimization problem reduces to minimizing:

$$\sum_{i=1}^{n}|y_i - \boldsymbol{w}^T\boldsymbol{x}_i|$$

This is the empirical risk minimization objective with the L1 loss function.

3. **Bayesian Optimization** (5 points each) Check the GP notebooks.