EEcs 553 HW8

Lingqi Huang

November 2024

1 Problem 1

(a): To avoid confusion, I will still use the original notation. We first notice that :

$$P(y=k|\mathbf{x},\pi^{'}) = \frac{P(\mathbf{x}|y=k,\pi^{'})P(y=k|\pi^{'})}{P(\mathbf{x}|\pi^{'})} \tag{1}$$
$$= \frac{P(\mathbf{x}|y=k,\pi)\pi_k^{'}}{P(\mathbf{x}|\pi^{'})} \tag{2}$$

because we know that $\phi_k(\mathbf{x})$ is unaffected by the shift. Now we also know that

$$P(\mathbf{x}|y=k,\pi) = \frac{P(y=k|\mathbf{x},\pi)P(\mathbf{x}|\pi)}{\pi_k}$$

Now if we substitute this term into (1), we will observe that

$$P(y=k|\mathbf{x},\pi^{'}) = P(y=k|\mathbf{x},\pi)\cdot\frac{\pi_k^{'}}{\pi_k}\cdot\frac{P(\mathbf{x}|\pi)}{P(\mathbf{x}|\pi^{'})}$$

Observe that

$$P(\mathbf{x}|\pi^{'}) = \sum_{j=0}^{1} P(\mathbf{x}|y=j,\pi^{'})\pi_j^{'} \tag{3}$$
$$= \sum_{j=0}^{1} P(\mathbf{x}|y=j,\pi)\pi_j^{'} \quad \text{(because } \phi_k(\mathbf{x}) \text{ is unaffected by shift)} \tag{4}$$
$$= P(\mathbf{x}|\pi)\sum_{j=0}^{1}\frac{P(y=j|\mathbf{x},\pi)\pi_j^{'}}{\pi_j} \tag{5}$$

Therefore, we finally conclude that

$$P(y=k|\mathbf{x},\pi^{'}) = \frac{\frac{\pi_k^{'}}{\pi_k}P(y=k|\mathbf{x},\pi)}{\frac{\pi_0^{'}}{\pi_0}P(y=0|\mathbf{x},\pi)+\frac{\pi_1^{'}}{\pi_1}P(y=1|\mathbf{x},\pi)}$$

which is exactly the same expression if we use the shorthand notation. This completes the proof.
(b) & (c): For the incomplete data likelihood function, we have that

$$l(\pi^{'}) = \log p(\mathbb{X}|\pi^{'}) \tag{6}$$
$$= \log\prod_{i=1}^{m}[\phi_1(\mathbf{x}_i)\pi^{'} + \phi_0(\mathbf{x}_i)(1-\pi^{'})] \tag{7}$$
$$= \sum_{i=1}^{m}\log\left(\phi_1(\mathbf{x}_i)\pi^{'} + \phi_0(\mathbf{x}_i)(1-\pi^{'})\right) \tag{8}$$

Now for the complete data log likelihood function, we have that

$$\log P(\mathbb{X}, \mathbb{Z}|\pi^{'}) = \log P(\mathbb{Z}|\mathbb{X}, \pi^{'}) + \log P(\mathbb{X}|\pi^{'})$$

Now we need to find $\log P(\mathbb{Z}|\mathbb{X}, \pi^{'})$, that is :

$$\log P(\mathbb{Z}|\mathbb{X}, \pi^{'}) = \sum_{i=1}^{m} y_i \log P(y_i = 1|\mathbb{X}, \pi^{'}) + (1 - y_i) \log P(y_i = 0|\mathbb{X}, \pi^{'})$$

Thus, combine the $\log P(\mathbb{Z}|\mathbb{X}, \pi^{'})$ and $\log P(\mathbb{X}|\pi^{'})$ will give us the full log-likelihood.
E-Step
Now to compute $Q(\pi^{'}; \pi^{'(t)})$, we observe that we need to compute that

$$E(y_i|\mathbb{X}, \pi^{'(t)}) = P^{(t)}(y = 1|\mathbb{X}, \pi^{'(t)})$$

that denote the randomness of $Q(\pi; \pi^{'(t)})$. We also notice that by Bayes rule, we have that

$$P(y_i = 1|\mathbb{X}, \pi^{'}) = \frac{\phi_1(\mathbf{x}_i)\pi^{'}}{(1 - \pi^{'})\phi_0(\mathbf{x}_i) + \pi^{'}\phi_1(\mathbf{x}_i)}$$

Thus, plug the above term into $Q(\pi^{'}; \pi^{'(t)})$ and after some simplification, we find that

$$\begin{aligned} Q(\pi^{'}; \pi^{'(t)}) = \sum_{i=1}^{m} & P^{(t)}(y_i = 1|\mathbb{X}, \pi^{'(t)}) \cdot [\log \phi_1(\mathbf{x}_i) + \log \pi^{'} - \log((1 - \pi^{'})\phi_0(\mathbf{x}_i) + \pi^{'}\phi_1(\mathbf{x}_i))] & (9) \\ & + (1 - P^{(t)}(y_i = 1|\mathbb{X}, \pi^{'(t)})) \cdot [\log \phi_0(\mathbf{x}_i) + \log(1 - \pi^{'}) - \log((1 - \pi^{'})\phi_0(\mathbf{x}_i) + \pi^{'}\phi_1(\mathbf{x}_i))] & (10) \\ & + \sum_{i=1}^{m} \log \left(\phi_1(\mathbf{x}_i)\pi^{'} + \phi_0(\mathbf{x}_i)(1 - \pi^{'})\right) & (11) \end{aligned}$$

M-Step
Now take derivative respect to $\pi^{'}$, we will finally simplify to the below expression that:

$$\frac{\partial Q}{\partial \pi^{'}} = \frac{1}{\pi^{'}(1 - \pi^{'})} \sum_{i=1}^{m} P^{(t)}(y_i = 1|\mathbf{x}_i, \pi^{'(t)}) - \frac{m}{1 - \pi^{'}} = 0$$

Thus, we can then conclude that the final M-step update reduces to:

$$\pi^{'(t+1)} = \frac{1}{m} \sum_{i=1}^{m} P^{(t)}(y_i = 1|\mathbf{x}_i, \pi^{'(t)})$$

where $P^{(t)}(y_i = 1|\mathbf{x}_i, \pi^{'(t)})$ is obtained after we replacing $\pi^{'}_k$ with $\pi^{'(t)}$ and $\mathbf{x}$ with $\mathbf{x}_i$. This completes part(b) and part(c).
(d) Please check the py and ipynb file I uploaded on canvas to see the code, or you can see it in the end of the file. The unadjust LR has accuracy of 0.83, the EM-adjusted LR has accuracy of 0.9, the accuracy of Clairvoyant adjusted LR is 0.9.

2 Problem 2

Given that $\epsilon_i \sim Laplacian(\beta)$, we conclude that $y_i \sim Laplacian(w^T x_i, \beta)$ that is

$$p(y_i) = \frac{\beta}{2} \exp(-\beta|y_i - w^T x_i|)$$

Thus, the likelihood can be written as

$$\begin{aligned} L(\mathbf{y}|w) &= \prod_{i=1}^{n} \frac{\beta}{2} \exp(-\beta|y_i - w^T x_i|) & (12) \\ &= \left(\frac{\beta}{2}\right)^n \exp\left(-\beta \sum_{i=1}^{n} |y_i - w^T x_i|\right) & (13) \end{aligned}$$

so the log-likelihood could be written as

$$l(w|\mathbf{y}) = -\beta \sum_{i=1}^{n} |y_i - w^T x_i| + n \log\left(\frac{\beta}{2}\right)$$

Thus, to maximize the likelihood and find w to achieve this, it is same to solve the minimization problem that

$$\min_{w} \frac{1}{n} \sum_{i=1}^{n} |y_i - w^T x_i| - \frac{1}{\beta} \log\left(\frac{\beta}{2}\right)$$

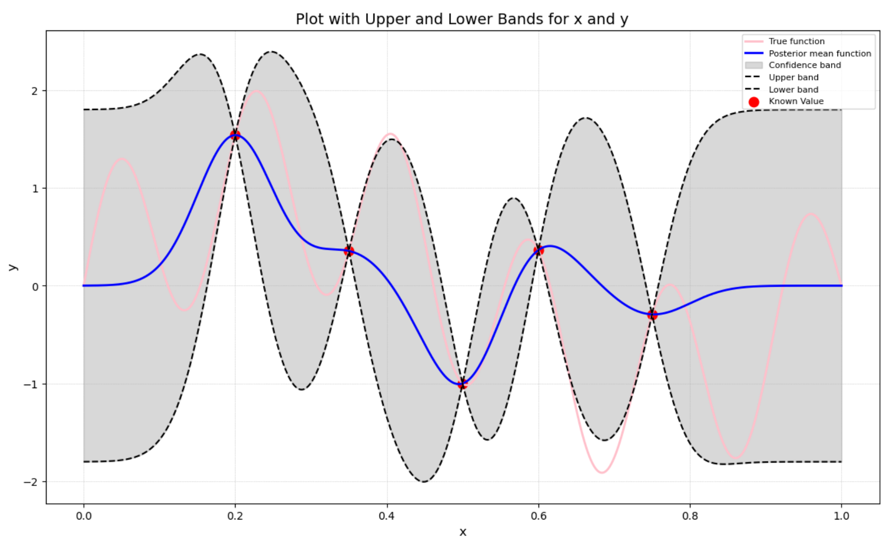
and notice the last term is unrelated to w, therefore, it is same to solve problem of empirical risk minimization that

$$\min_{w} \frac{1}{n} \sum_{i=1}^{n} L(y_i, f(x_i))$$

that the loss function is $L(y,t) = |y - t|$ and $f(x_i) = w^T x_i$. This completes the proof.

3 Problem 3

(a): please check the below picture.



(b): Please check the code I uploaded in canvas, or at the end of the page. The next value of x that going to evaluate is 0.2474.

(c): The next point that PI method suggest to evaluate is 0.2004.

(d): The next point that EI method suggest to evaluate is 0.234. Notice here I did not use the exact formula for calculating expectation, instead, I sample points from the normal distribution and then average them to get the EI for each point, which make sense by law of large number.(Monte-Carlo)

(e): All three method will approximately converge to 0.2276 with function value of 1.99. The detail of algorithm, number of iteration, epsilon can be seen in the below code I attached.