CIS 520, Machine Learning, Fall 2018: Assignment 8

Shubhankar Patankar

Collaborator:

Simran Arora

1 Active Learning

For this problem, I referenced Professor Shivani Agarwal's lecture notes on Semi-Supervised and Active Learning.

1.1 Part 1

1. x_1 is selected. With a $0-1$ loss function, we determine that our model has lower performance with more misclassifications. In uncertainty sampling, we query the instance for which the algorithm is currently most uncertain regarding its label. With the $0-1$ loss, the algorithm is uncertain about labels of instances x for which the estimated probability of a positive label $\hat{p}(+1|x)$ is close to the decision threshold of $\frac{1}{2}$. Thus of the three given instances, the uncertainty sampling method would select x_1 because $\iota(x_1)$ is closest to $\frac{1}{2}$ and since the model performs binary classification, we know all the probabilities of x_i and $\bar{x_i}$.

2. For a cost-sensitive loss function, we can similarly select an instance for which the estimated probability of a positive label is closest to the appropriate decision threshold. In the $0-1$ loss, the decision boundary is $\frac{1}{2}$. To determine the decision boundary for this cost-sensitive case, we want to determine the cost of deciding to predict that x is $+1$ against the cost of deciding to predict that x is -1. The cost to predict -1 is:

$$Cost(\hat{y} = -1,\, y = -1)P(y_i = -1|x) + Cost(\hat{y} = -1,\, y = +1)P(y_i = +1|x)$$

   and the cost to predict $+1$ is:

$$Cost(\hat{y} = +1,\, y = +1)P(y_i = +1|x) + Cost(\hat{y} = -1,\, y = +1)P(y_i = +1|x)$$

   For x_1, the negative prediction evaluates to $(0)(0.45) + (0.2)(0.45) = 0.09$ and the positive prediction evaluates to $(0)(0.55) + (0.8)(0.55) = 0.44$. So the total cost for $x_1 = 0.09 + 0.44 = 0.53$. For x_2, the negative prediction evaluates to $(0)(0.05) + (0.2)(0.05) = 0.01$ and the positive prediction evaluates to $(0)(0.95) + (0.8)(0.95) = 0.76$. So the total cost for $x_2 = 0.01 + 0.76 = 0.77$. For x_3, the negative prediction evaluates to $(0)(0.7) + (0.2)(0.7) = 0.35$ and the positive prediction evaluates to $(0)(0.3) + (0.8)(0.3) = 0.24$. So the total cost for $x_3 = 0.35 + 0.24 = 0.59$. Thus in this case, x_2 has the highest cost so we should choose x_2.

1.2 Part 2

1. We select x_2. Using the uncertainty sampling approach based on least confident prediction, we view the estimated probability of the predicted label as a measure of confidence or 'certainty' and select an instance for which the predicted label comes with least confidence. We want to determine

$$x^* \in argmin_x(max_y \hat{p}(y|x))$$

For x_1, we have that $max_y \hat{p}(y|x) = 1$ because the $\iota_1(x_1) = 0.8$ for label 1, which is the highest probability we have for a label given x_1. For x_2, we have that $max_y \hat{p}(y|x) = 3$ because the $\iota_3(x_2) = 0.45$ for label 3, which is the highest probability we have for a label given x_2. For x_3, we have that $max_y \hat{p}(y|x) = 3$ because the $\iota_3(x_3) = 0.5$ for label 13, which is the highest probability we have for a label given x_3. Next, of these values, we want to take the $argmin_x$, and thus we select x_2 because $0.45 < 0.5$ and $0.45 < 0.8$.

2. Using the uncertainty sampling approach based on the margin between the two highest-probability predictions, we want to pick an instance for which the algorithm is most 'confused' between two labels, i.e. has the smallest margen or gap between the probabilities of the top two predicted labels. If $\hat{y}(x)$ is equal to the label with the highest estimated probability for x:

$$\hat{y}(x) \in argmax_y \hat{p}(y|x)$$

We have the following values for $\hat{y}(x)$:

- $\hat{y}(x_1) = 1$ as we discussed in 1.
- $\hat{y}(x_2) = 3$ as we discussed in 1.
- $\hat{y}(x_3) = 3$ as we discussed in 1.

We can evaluate for each x_i the value of $(max_{y \neq \hat{y}(x)} \hat{p}(y|x))$:

- For x_1, $(max_{y \neq \hat{y}(x)} \hat{p}(y|x)) = 0.15$ because label 3 has the next highest probability after label 1, which is 0.15.
- For x_2, $(max_{y \neq \hat{y}(x)} \hat{p}(y|x)) = 0.35$ because label 2 has the next highest probability after label 3 with probability 0.35.
- For x_3, $(max_{y \neq \hat{y}(x)} \hat{p}(y|x)) = 0.45$ because label 2 has the next highest probability after label 3 with probability 0.45.

Now given this, we select an instance according to: $x^* \in argmin_x(max_y \hat{p}(y|x) - max_{y \neq \hat{y}(x)} \hat{p}(y|x))$

- For x_1 we have $0.8 - 0.15 = 0.65$.
- For x_2 we have $0.45 - 0.35 = 0.10$.
- For x_3 we have $0.5 - 0.45 = 0.05$.

We can see that the x_i that minimizes this difference is x_3 with difference 0.05 and thus we would select x_3 according to this method.

2 Reinforcement Learning

1. The MDP is formulated as a tuple $(\mathbf{S}, \mathbf{A}, p, r, \gamma)$, where:

- $\mathbf{S}$ is the set of states in our environment.
- $\mathbf{A}$ is the set of actions.
- p is the function that takes in the current state and action as inputs, and outputs the probabilities for each of the next states at which the agent could end up.
- r is the reward function that takes in a given initial state s, action taken a, and final state s' and determines the reward to give the agent.
- γ is the discount factor (here $\gamma = 0.9$).

$$\mathbf{S} \in \{0, 1, 2, 3, 4\}$$

$$\mathbf{A} \in \{forward, backward\}$$

p is defined based on the following known transition probabilities between states with the associated transition rewards r:

- $s = 0$:
  - $p(s' = 0|s = 0, a = forward) = 0.1$
    $r(s = 0, a = forward, s' = 0) = 2$
  - $p(s' = 1|s = 0, a = forward) = 0.9$
    $r(s = 0, a = forward, s' = 1) = 0$
  - $p(s' = 2|s = 0, a = forward) = 0$
    $r(s = 0, a = forward, s' = 2) = 0$
  - $p(s' = 3|s = 0, a = forward) = 0$
    $r(s = 0, a = forward, s' = 3) = 0$
  - $p(s' = 4|s = 0, a = forward) = 0$
    $r(s = 0, a = forward, s' = 4) = 0$
  - $p(s' = 0|s = 0, a = backward) = 0.9$
    $r(s = 0, a = backward, s' = 0) = 2$
  - $p(s' = 1|s = 0, a = backward) = 0.1$
    $r(s = 0, a = backward, s' = 1) = 0$
  - $p(s' = 2|s = 0, a = backward) = 0$
    $r(s = 0, a = backward, s' = 2) = 0$
  - $p(s' = 3|s = 0, a = backward) = 0$
    $r(s = 0, a = backward, s' = 3) = 0$
  - $p(s' = 4|s = 0, a = backward) = 0$
    $r(s = 0, a = backward, s' = 4) = 0$
- $s = 1$:
  - $p(s' = 0|s = 1, a = forward) = 0.1$
    $r(s = 1, a = forward, s' = 0) = 2$
  - $p(s' = 1|s = 1, a = forward) = 0$
    $r(s = 1, a = forward, s' = 1) = 0$
  - $p(s' = 2|s = 1, a = forward) = 0.9$
    $r(s = 1, a = forward, s' = 2) = 0$
  - $p(s' = 3|s = 1, a = forward) = 0$
    $r(s = 1, a = forward, s' = 3) = 0$
  - $p(s' = 4|s = 1, a = forward) = 0$
    $r(s = 1, a = forward, s' = 4) = 0$
  - $p(s' = 0|s = 1, a = backward) = 0.9$
    $r(s = 1, a = backward, s' = 0) = 2$
  - $p(s' = 1|s = 1, a = backward) = 0$
    $r(s = 1, a = backward, s' = 1) = 0$
  - $p(s' = 2|s = 1, a = backward) = 0.1$
    $r(s = 1, a = backward, s' = 2) = 0$
  - $p(s' = 3|s = 1, a = backward) = 0$
    $r(s = 1, a = backward, s' = 3) = 0$
  - $p(s' = 4|s = 1, a = backward) = 0$
    $r(s = 1, a = backward, s' = 4) = 0$
- $s = 2$:
  - $p(s' = 0|s = 2, a = forward) = 0.1$
    $r(s = 2, a = forward, s' = 0) = 2$
  - $p(s' = 1|s = 2, a = forward) = 0$
    $r(s = 2, a = forward, s' = 1) = 0$
  - $p(s' = 2|s = 2, a = forward) = 0$
    $r(s = 2, a = forward, s' = 2) = 0$

- $p(s' = 3|s = 2, a = forward) = 0.9$
  $r(s = 2, a = forward, s' = 3) = 0$
- $p(s' = 4|s = 2, a = forward) = 0$
  $r(s = 2, a = forward, s' = 4) = 0$
- $p(s' = 0|s = 2, a = backward) = 0.9$
  $r(s = 2, a = backward, s' = 0) = 2$
- $p(s' = 1|s = 2, a = backward) = 0$
  $r(s = 2, a = backward, s' = 1) = 0$
- $p(s' = 2|s = 2, a = backward) = 0$
  $r(s = 2, a = backward, s' = 2) = 0$
- $p(s' = 3|s = 2, a = backward) = 0.1$
  $r(s = 2, a = backward, s' = 3) = 0$
- $p(s' = 4|s = 2, a = backward) = 0$
  $r(s = 2, a = backward, s' = 4) = 0$

- $s = 3$:
  - $p(s' = 0|s = 3, a = forward) = 0.1$
    $r(s = 3, a = forward, s' = 0) = 2$
  - $p(s' = 1|s = 3, a = forward) = 0$
    $r(s = 3, a = forward, s' = 1) = 0$
  - $p(s' = 2|s = 3, a = forward) = 0$
    $r(s = 3, a = forward, s' = 2) = 0$
  - $p(s' = 3|s = 3, a = forward) = 0$
    $r(s = 3, a = forward, s' = 3) = 0$
  - $p(s' = 4|s = 3, a = forward) = 0.9$
    $r(s = 3, a = forward, s' = 4) = 0$
  - $p(s' = 0|s = 3, a = backward) = 0.9$
    $r(s = 3, a = backward, s' = 0) = 2$
  - $p(s' = 1|s = 3, a = backward) = 0$
    $r(s = 3, a = backward, s' = 1) = 0$
  - $p(s' = 2|s = 3, a = backward) = 0$
    $r(s = 3, a = backward, s' = 2) = 0$
  - $p(s' = 3|s = 3, a = backward) = 0$
    $r(s = 3, a = backward, s' = 3) = 0$
  - $p(s' = 4|s = 3, a = backward) = 0.1$
    $r(s = 3, a = backward, s' = 4) = 0$

- $s = 4$:
  - $p(s' = 0|s = 4, a = forward) = 0.1$
    $r(s = 4, a = forward, s' = 0) = 2$
  - $p(s' = 1|s = 4, a = forward) = 0$
    $r(s = 4, a = forward, s' = 1) = 0$
  - $p(s' = 2|s = 4, a = forward) = 0$
    $r(s = 4, a = forward, s' = 2) = 0$
  - $p(s' = 3|s = 4, a = forward) = 0$
    $r(s = 4, a = forward, s' = 3) = 0$
  - $p(s' = 4|s = 4, a = forward) = 0.1$
    $r(s = 4, a = forward, s' = 4) = 10$
  - $p(s' = 0|s = 4, a = backward) = 0.9$
    $r(s = 4, a = backward, s' = 0) = 2$
  - $p(s' = 1|s = 4, a = backward) = 0$
    $r(s = 4, a = backward, s' = 1) = 0$

- $p(s' = 2|s = 4, a = backward) = 0$
  $r(s = 4, a = backward, s' = 2) = 0$
- $p(s' = 3|s = 4, a = backward) = 0$
  $r(s = 4, a = backward, s' = 3) = 0$
- $p(s' = 4|s = 4, a = backward) = 0.1$
  $r(s = 4, a = backward, s' = 4) = 10$

Transition probabilities and associated rewards are summarized in the following matrices:

$$p(s'|a = forward, s) = \begin{bmatrix} 0.1 & 0.9 & 0 & 0 & 0 \\ 0.1 & 0 & 0.9 & 0 & 0 \\ 0.1 & 0 & 0 & 0.9 & 0 \\ 0.1 & 0 & 0 & 0 & 0.9 \\ 0.1 & 0 & 0 & 0 & 0.9 \end{bmatrix}$$

$$r(s, a = forward, s') = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 10 \end{bmatrix}$$

$$p(s'|a = backward, s) = \begin{bmatrix} 0.9 & 0.1 & 0 & 0 & 0 \\ 0.9 & 0 & 0.1 & 0 & 0 \\ 0.9 & 0 & 0 & 0.1 & 0 \\ 0.9 & 0 & 0 & 0 & 0.1 \\ 0.9 & 0 & 0 & 0 & 0.1 \end{bmatrix}$$

$$r(s, a = backward, s') = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 10 \end{bmatrix}$$

$$[p(s'|a, s)]_{ij} = probability\ of\ ending\ up\ in\ state\ j\ when\ taking\ action\ a\ in\ state\ i$$

$$[r(s, a, s')]_{ij} = reward\ earned\ when\ moving\ from\ state\ i\ to\ j\ under\ action\ a$$

2. Optimal State Values

$$V^*(s) = \{40.7420, 45.5250, 51.4299, 58.7199, 67.7199\}$$

3. Optimal Policy in States

$$\pi^*(s) = \{forward, forward, forward, forward, forward\}$$

3 Semi-Supervised Learning

1. We want to determine the initial parameters ($\hat{\Theta}^0_1$, $\hat{\Theta}^0_{1|+1}$, $\hat{\Theta}^0_{1|-1}$, $\hat{\Theta}^0_{2|+1}$, $\hat{\Theta}^0_{2|-1}$):

   - For $\hat{\theta}^0_1 = \frac{4}{8} = \frac{1}{2}$, we have $Y = +1$ in 4 of 8 labeled instances.
   - For $\hat{\theta}^0_{1|+1} = \frac{3}{4}$ we have $Y = +1$ n 4 of 8 labeled instances. We then have $\vec{x}$ of the form$(1, *)$ in 3 of those 4 instances.
   - For $\hat{\theta}^0_{1|-1} = \frac{1}{4}$ we have $Y = -1$ in 4 of 8 labeled instances. We then have $\vec{x}$ of the form$(1, *)$ in 1 of those 4 instances.

- For $\hat{\theta}^0_{2|+1} = \frac{1}{2}$ we have $Y = +1$ in 4 of 8 labeled instances. We then have $\vec{x}$ of the form$(*, 1)$ in 2 of those 4 instances.
- For $\hat{\theta}^0_{2|-1} = \frac{1}{2}$ we have $Y = -1$ in 4 of 8 labeled instances. We then have $\vec{x}$ of the form$(1, *)$ in 3 of those 4 instances.

2. We first want to determine $q^0(+1|\vec{x} = (1,\ 1)) = P(Y = +1|\vec{x} = (1,\ 1); \hat{\theta}^0_1)$. Using Bayes' Rule, we can rewrite this as the following:

$$P(Y = +1|\vec{x} = (1,\ 1); \hat{\theta}^0_1) = \frac{P(Y = +1 \cap \vec{x} = (1,\ 1)}{P(\vec{x} = (1,\ 1))}$$

Applying Bayes' Rule again, we get the following:

$$P(Y = +1|\vec{x} = (1,\ 1); \hat{\theta}^0_1) = \frac{P(Y = +1)P(\vec{x} = (1,\ 1)|Y = +1)}{P(\vec{x} = (1,\ 1))}$$

In the part 1, we see that $\hat{\theta}^0_1 = P(Y = +1) = \frac{1}{2}$. By the example given in the problem statement, we know we can express $P(\vec{x} = (1,\ 1)|Y = +1)$ as $(\hat{\theta}^0_{1|+1})(\theta^0_{2|+1}) = (\frac{3}{4})(\frac{1}{2})$. Finally in the denominator we can apply the total probability formula to express $P(\vec{x} = (1,\ 1))$ as $P(\vec{x} = (1,\ 1)|Y = +1)P(Y = +1) + P(\vec{x} = (1,\ 1)|Y = -1)P(Y = -1) = [(\frac{1}{2})(\frac{3}{4})(\frac{1}{2}) + (\frac{1}{2})(\frac{1}{4})(\frac{1}{2})]$. Thus, the value of $q^0(+1|\vec{x} = (1,\ 1)$ is:

$$\frac{(\frac{1}{2})(\frac{3}{4})(\frac{1}{2})}{[(\frac{1}{2})(\frac{3}{4})(\frac{1}{2}) + (\frac{1}{2})(\frac{1}{4})(\frac{1}{2})]} = \mathbf{\frac{3}{4}}$$

Next we want to determine $q^0(+1|\vec{x} = (0,\ 0)) = P(Y = +1|\vec{x} = (0,\ 0); \hat{\theta}^0_1)$. Using Bayes' Rule, we can rewrite this as the following:

$$P(Y = +1|\vec{x} = (0,\ 0); \hat{\theta}^0_1) = \frac{P(Y = +1 \cap \vec{x} = (0,\ 0)}{P(\vec{x} = (0,\ 0))}$$

Applying Bayes' Rule again, we get the following:

$$P(Y = +1|\vec{x} = (0,\ 0); \hat{\theta}^0_1) = \frac{P(Y = +1)P(\vec{x} = (0,\ 0)|Y = +1)}{P(\vec{x} = (0,\ 0))}$$

In the part 1, we see that $\hat{\theta}^0_t = P(Y = +1) = \frac{1}{2}$. By the example given in the problem statement, we know we can express $P(\vec{x} = (0,\ 0)|Y = +1)$ as $(1 - \hat{\theta}^0_{1|+1})(1 - \theta^0_{2|+1}) = (1 - \frac{3}{4})(1 - \frac{1}{2}) = (\frac{1}{4})(\frac{1}{2})$. Finally in the denominator we can apply the total probability formula to express $P(\vec{x} = (1,\ 1))$ as $P(\vec{x} = (0,\ 0)|Y = +1)P(Y = +1) + P(\vec{x} = (0,\ 0)|Y = -1)P(Y = -1) = [(\frac{1}{2})(\frac{3}{4})(\frac{1}{2}) + (\frac{1}{2})(\frac{1}{4})(\frac{1}{2})]$. Thus, the value of $q^0(+1|\vec{x} = (0,\ 0))$ is:

$$\frac{(\frac{1}{2})(\frac{1}{4})(\frac{1}{2})}{[(\frac{1}{2})(\frac{3}{4})(\frac{1}{2}) + (\frac{1}{2})(\frac{1}{4})(\frac{1}{2})]} = \mathbf{\frac{1}{4}}$$

3. We want to determine the updated parameters $(\hat{\theta}^1_t,\ \hat{\theta}^1_{1|+1}, \hat{\theta}^1_{1|-1},\ \hat{\theta}^1_{2|+1},\ \hat{\theta}^1_{2|-1})$ and to do this we can use the update rules given in the problem statement in accordance with the EM formula. For these problems, we know that $m_L = 8$ and $m_U = 4$.

- For $\hat{\theta}^1_t$, we have the following:

$$\hat{\theta}^1_t = \frac{1}{8+4}(\sum_{i=1}^{8} \mathbf{1}(Y_i = +1) + \sum_{i=9}^{12} q^0(+1|\vec{x_i}))$$

$$= \frac{1}{12}(4 + [\frac{3}{4} + \frac{3}{4} + \frac{1}{4} + \frac{1}{4}] = \frac{6}{12} = \frac{1}{2}$$

- For $\hat{\theta}^1_{1|+1}$ we have the following:

$$\hat{\theta}^1_{1|+1} = \frac{\sum_{i=1}^{8} \mathbf{1}(Y_i = +1, X_{ij} = 1) + \sum_{i=9}^{12} q^0(+1|\vec{x_i})\mathbf{1}(x_{ij} = 1)}{\sum_{i=1}^{8} \mathbf{1}(Y_i = +1) + \sum_{i=9}^{12} q^0(+1|\vec{x_i})}$$

$$= \frac{3 + [\frac{3}{4} + \frac{3}{4}]}{4 + [\frac{3}{4} + \frac{3}{4} + \frac{1}{4} + \frac{1}{4}]} = \frac{4.5}{6} = \frac{3}{4}$$

- For $\hat{\theta}^1_{1|-1}$ we have the following:

$$\hat{\theta}^1_{1|-1} = \frac{\sum_{i=1}^{8} \mathbf{1}(Y_i = -1, X_{ij} = 1) + \sum_{i=9}^{12} q^0(-1|\vec{x_i})\mathbf{1}(x_{ij} = 1)}{\sum_{i=1}^{8} \mathbf{1}(Y_i = -1) + \sum_{i=9}^{12} q^0(-1|\vec{x_i})}$$

$$= \frac{1 + [\frac{1}{4} + \frac{1}{4}]}{4 + [\frac{3}{4} + \frac{3}{4} + \frac{1}{4} + \frac{1}{4}]} = \frac{1.5}{6} = \frac{1}{4}$$

- For $\hat{\theta}^1_{2|+1}$ we have the following:

$$\hat{\theta}^1_{2|+1} = \frac{\sum_{i=1}^{8} \mathbf{1}(Y_i = +1, X_{ij} = 1) + \sum_{i=9}^{12} q^0(+1|\vec{x_i})\mathbf{1}(x_{ij} = 1)}{\sum_{i=1}^{8} \mathbf{1}(Y_i = +1) + \sum_{i=9}^{12} q^0(+1|\vec{x_i})}$$

$$= \frac{2 + [\frac{3}{4} + \frac{3}{4}]}{4 + [\frac{3}{4} + \frac{3}{4} + \frac{1}{4} + \frac{1}{4}]} = \frac{3.5}{6} = \frac{7}{12}$$

- For $\hat{\theta}^1_{2|-1}$ we have the following:

$$\hat{\theta}^1_{2|-1} = \frac{\sum_{i=1}^{8} \mathbf{1}(Y_i = -1, X_{ij} = 1) + \sum_{i=9}^{12} q^0(-1|\vec{x_i})\mathbf{1}(x_{ij} = 1)}{\sum_{i=1}^{8} \mathbf{1}(Y_i = -1) + \sum_{i=9}^{12} q^0(-1|\vec{x_i})}$$

$$= \frac{2 + [\frac{1}{4} + \frac{1}{4}]}{4 + [\frac{3}{4} + \frac{3}{4} + \frac{1}{4} + \frac{1}{4}]} = \frac{2.5}{6} = \frac{5}{12}$$

4. Log-Likelihood

$$p(\mathbf{x}_i, y_i; \hat{\boldsymbol{\theta}}^t) = \left[p(y_i)\prod_{j=1}^{2} p(x_j\ |y_i)\right] = \left[p(y_i) \times p(x_1 \mid y_i) \times p(x_2 \mid y_i)\right]$$
$$= \left[(\hat{\theta}^t_{+1})(\hat{\theta}^t_{1|+1})^{x_{i1}}(1 - \hat{\theta}^t_{1|+1})^{1-x_{i1}}(\hat{\theta}^t_{2|+1})^{x_{i2}}(1 - \hat{\theta}^t_{2|+1})^{1-x_{i2}}\right]^{\mathbf{1}(y_i=+1)}$$
$$\times \left[(1 - \hat{\theta}^t_{+1})(\hat{\theta}^t_{1|-1})^{x_{i1}}(1 - \hat{\theta}^t_{1|-1})^{1-x_{i1}}(\hat{\theta}^t_{2|-1})^{x_{i2}}(1 - \hat{\theta}^t_{2|-1})^{1-x_{i2}}\right]^{\mathbf{1}(y_i=-1)}$$

$$p(\mathbf{x}_i; \hat{\boldsymbol{\theta}}^t) = \left[(\hat{\theta}^t_{+1})(\hat{\theta}^t_{1|+1})^{x_{i1}}(1 - \hat{\theta}^t_{1|+1})^{1-x_{i1}}(\hat{\theta}^t_{2|+1})^{x_{i2}}(1 - \hat{\theta}^t_{2|+1})^{1-x_{i2}}\right]$$
$$+ \left[(1 - \hat{\theta}^t_{+1})(\hat{\theta}^t_{1|-1})^{x_{i1}}(1 - \hat{\theta}^t_{1|-1})^{1-x_{i1}}(\hat{\theta}^t_{2|-1})^{x_{i2}}(1 - \hat{\theta}^t_{2|-1})^{1-x_{i2}}\right]$$

The log-likelihood is then written as follows:

$$ln\ p(S; \hat{\boldsymbol{\theta}}^t) = \sum_{i=1}^{m_L} ln\ p(\mathbf{x}_i, y_i; \hat{\boldsymbol{\theta}}^t) + \sum_{i=m_L+1}^{m_L+m_U} ln\ p(\mathbf{x}_i; \hat{\boldsymbol{\theta}}^t)$$

where, $p(\mathbf{x}_i, y_i; \hat{\boldsymbol{\theta}}^t)$ and $p(\mathbf{x}_i; \hat{\boldsymbol{\theta}}^t)$ are defined above.
For the given dataset for the labelled points, $ln\ p(\mathbf{x}_i, y_i; \hat{\boldsymbol{\theta}}^t)$ evaluates to:

For $\mathbf{x} = (1, 1)$ and $y = +1$: (contributes two such terms)

$$2\ ln\ p(\mathbf{x}_i, y_i; \hat{\boldsymbol{\theta}}^t) = 2\ ln\ \left[(\hat{\theta}^t_{+1})(\hat{\theta}^t_{1|+1})(\hat{\theta}^t_{2|+1})\right]$$

For $\mathbf{x} = (1, 0)$ and $y = +1$:

$$ln\ p(\mathbf{x}_i, y_i; \hat{\boldsymbol{\theta}}^t) = ln\ \left[(\hat{\theta}^t_{+1})(\hat{\theta}^t_{1|+1})(1 - \hat{\theta}^t_{2|+1})\right]$$

For $\mathbf{x} = (0, 0)$ and $y = +1$:

$$ln\ p(\mathbf{x}_i, y_i; \hat{\boldsymbol{\theta}}^t) = ln\ \left[(\hat{\theta}^t_{+1})(1 - \hat{\theta}^t_{1|+1})(1 - \hat{\theta}^t_{2|+1})\right]$$

For $\mathbf{x} = (1, 0)$ and $y = -1$:

$$ln\ p(\mathbf{x}_i, y_i; \hat{\boldsymbol{\theta}}^t) = ln\ \left[(1 - \hat{\theta}^t_{+1})(\hat{\theta}^t_{1|-1})(1 - \hat{\theta}^t_{2|-1})\right]$$

For $\mathbf{x} = (0, 1)$ and $y = -1$: (contributes two such terms)

$$2\ ln\ p(\mathbf{x}_i, y_i; \hat{\boldsymbol{\theta}}^t) = 2\ ln\ \left[(1 - \hat{\theta}^t_{+1})(1 - \hat{\theta}^t_{1|-1})(\hat{\theta}^t_{2|-1})\right]$$

For $\mathbf{x} = (0, 0)$ and $y = -1$:

$$ln\ p(\mathbf{x}_i, y_i; \hat{\boldsymbol{\theta}}^t) = ln\ \left[(1 - \hat{\theta}^t_{+1})(1 - \hat{\theta}^t_{1|-1})(1 - \hat{\theta}^t_{2|-1})\right]$$

For the given dataset for the unlabelled points, $ln\ p(\mathbf{x}_i, ; \hat{\boldsymbol{\theta}}^t)$ evaluates to:

For $\mathbf{x} = (1, 1)$: (contributes two such terms)

$$2\ ln\ \left[(\hat{\theta}^t_{+1})(\hat{\theta}^t_{1|+1})(\hat{\theta}^t_{2|+1}) + (1 - \hat{\theta}^t_{+1})(\hat{\theta}^t_{1|-1})(\hat{\theta}^t_{2|-1})\right]$$

For $\mathbf{x} = (0, 0)$: (contributes two such terms)

$$2\ ln\ \left[(\hat{\theta}^t_{+1})(1 - \hat{\theta}^t_{1|+1})(1 - \hat{\theta}^t_{2|+1}) + (1 - \hat{\theta}^t_{+1})(1 - \hat{\theta}^t_{1|-1})(1 - \hat{\theta}^t_{2|-1})\right]$$

The final log-likelihood $ln\ p(S; \hat{\boldsymbol{\theta}}^t)$ is the sum of all terms above.