CIS 520, Machine Learning, Fall 2018: Assignment 8

Wentao He (wentaoh)

Collaborator:N/A

1 Active Learning

1.1 Part 1

1. $\boxed{x_1}$. The final classifier would be evaluated in terms of 0-1 loss, and the loss function is:
   Expected Loss $= y_i(\hat{\eta}(x_i)C_{TP} + (1 - \hat{\eta}(x_i))C_{TN} + (1 - y_i)(\hat{\eta}(xi)C_{FP} + (1 - \hat{\eta}(x_i))C_{FN})$. Therefore:
   Expected Loss $(x_1) = C_{TP} \times \hat{\eta}(x_i) + C_{FP} \times (1 - \hat{\eta}(x_i)) = 0 \times 0.55 + 1 \times 0.45 = 0.45$
   Expected Loss $(x_2) = C_{TP} \times \hat{\eta}(x_i) + C_{FP} \times (1 - \hat{\eta}(x_i)) = 0 \times 0.95 + 1 \times 0.05 = 0.05$
   Expected Loss $(x_3) = C_{TN} \times \hat{\eta}(x_i) + C_{FN} \times (1 - \hat{\eta}(x_i)) = 1 \times 0.3 + 0 \times 0.7 = 0.3$
   And we know from the calculation that the label of x_1 will reduce the loss the most.

2. $\boxed{x_2}$. Similar to question 1, we get:
   Expected Loss $(x_1) = C_{TP} \times \hat{\eta}(x_i) + C_{FP} \times (1 - \hat{\eta}(x_i)) = 0 \times 0.55 + 0.8 \times 0.45 = 0.036$
   Expected Loss $(x_2) = C_{TP} \times \hat{\eta}(x_i) + C_{FP} \times (1 - \hat{\eta}(x_i)) = 0 \times 0.95 + 0.8 \times 0.05 = 0.04$
   Expected Loss $(x_3) = C_{TN} \times \hat{\eta}(x_i) + C_{FN} \times (1 - \hat{\eta}(x_i)) = 0.2 \times 0.3 + 0 \times 0.7 = 0.06$
   And we know from the calculation that the label of x_2 will reduce the loss the most.

1.2 Part 2

1. $\boxed{x_2}$. The classifier would pick the instance that has the least confidence in its most likely label based on the least confidence strategy. x_2 has the lowest confidence in its most likely label, which is 0.45.

2. $\boxed{x_3}$. The classifier would pick the instance that has the smallest difference between the first and second labels that are the most probable based on the margin sampling strategy. x_3 has the smallest difference.

2 Reinforcement Learning

1. MDP

   (a) $S = [0, 1, 2, 3, 4]$

   (b) $A = [f, b]$

   (c) For p:
   $p(1|0, f) = 0.9 \quad p(0|0, f) = 0.1 \quad p(1|0, b) = 0.1 \quad p(0|0, b) = 0.9$
   $p(0|1, f) = 0.1 \quad p(2|1, f) = 0.9 \quad p(0|1, b) = 0.9 \quad p(2|1, b) = 0.1$
   $p(0|2, f) = 0.1 \quad p(3|2, f) = 0.9 \quad p(0|2, b) = 0.9 \quad p(3|2, b) = 0.1$
   $p(0|3, f) = 0.1 \quad p(4|3, f) = 0.9 \quad p(0|3, b) = 0.9 \quad p(4|3, b) = 0.1$
   $p(0|4, f) = 0.1 \quad p(4|4, f) = 0.9 \quad p(0|4, b) = 0.9 \quad p(4|4, b) = 0.1$

   (d) For r:
   $r(0, f, 1) = 0 \quad r(0, f, 0) = 2 \quad r(0, b, 1) = 0 \quad r(0, b, 0) = 2$
   $r(0, f, 1) = 0 \quad r(0, f, 0) = 2 \quad r(0, b, 1) = 0 \quad r(0, b, 0) = 2$
   $r(2, f, 0) = 2 \quad r(2, f, 3) = 0 \quad r(2, b, 0) = 2 \quad r(2, b, 3) = 0$
   $r(3, f, 0) = 2 \quad r(3, f, 4) = 0 \quad r(3, b, 0) = 2 \quad r(3, b, 4) = 0$
   $r(4, f, 0) = 2 \quad r(4, f, 4) = 10 \quad r(4, b, 0) = 2 \quad r(4, b, 4) = 10$

2. Optimal State-Value Function
   $V^*(1) = 40.74 \quad V^*(2) = 45.53 \quad V^*(3) = 51.43 \quad V^*(4) = 58.72 \quad V^*(5) = 67.72;$

3. Optimal Deterministic Policy
   $\pi^*(s_i) = 1, \forall i \in [1, 2, 3, 4, 5]$

3 Semi-Supervised Learning

1. Initial Maximum Likelihood Parameters
   $\hat{\theta}^0_{+1} = \frac{1}{2}$ $\quad \hat{\theta}^0_{1|+1} = \frac{3}{4}$ $\quad \hat{\theta}^0_{2|+1} = \frac{1}{2}$ $\quad \hat{\theta}^0_{1|-1} = \frac{1}{4}$ $\quad \hat{\theta}^0_{2|-1} = \frac{1}{2}$

2. E Step

$$\begin{aligned}
q_0(+1|x=(1,1)) &= P(Y=+1|x=(1,1);\theta^0) \\
&= \frac{\hat{\theta}^0_{+1} \times \hat{\theta}^0_{1|+1} \times \hat{\theta}^0_{2|+1}}{\hat{\theta}^0_{+1} \times \hat{\theta}^0_{1|+1} \times \hat{\theta}^0_{2|+1} + \hat{\theta}^0_{-1} \times \hat{\theta}^0_{1|-1} \times \hat{\theta}^0_{2|-1}} \\
&= \frac{0.5 * 0.75 * 0.5}{0.5 * 0.75 * 0.5 + 0.5 * 0.25 * 0.5} \\
&= \frac{3}{4} \\
q_0(+1|x=(0,0)) &= P(Y=+1|x=(0,0);\theta^0) \\
&= \frac{\hat{\theta}^0_{+1} \times (1-\hat{\theta}^0_{1|+1}) \times (1-\hat{\theta}^0_{2|+1})}{\hat{\theta}^0_{+1} \times (1-\hat{\theta}^0_{1|+1}) \times (1-\hat{\theta}^0_{2|+1}) + \hat{\theta}^0_{-1} \times (1-\hat{\theta}^0_{1|-1}) \times (1-\hat{\theta}^0_{2|-1})} \\
&= \frac{0.5 * 0.25 * 0.5}{0.5 * 0.25 * 0.5 + 0.5 * 0.25 * 0.5} \\
&= \frac{1}{4}
\end{aligned}$$

3. M Step

$$\begin{aligned}
\hat{\theta}^0_{+1} &= \frac{1}{12}(\sum_{i=1}^{m_L} \mathbf{1}(y_i=+1) + \sum_{i=m_L+1}^{m_L+m_U} q^0(+1|x_i)) \\
&= \frac{1}{12} \times (4 + 2 \times 0.75 + 2 \times 0.25) \\
&= \frac{1}{2} \\
\hat{\theta}^0_{1|+1} &= \frac{\sum_{i=1}^{m_L} \mathbf{1}(y_i=1, x_{i,1}=1) + \sum_{i=m_L+1}^{m_L+m_U} q^1(+1|x_i)\mathbf{1}(x_{i,1}=1)}{\sum_{i=1}^{m_L} \mathbf{1}(y_i=+1) + \sum_{i=m_L+1}^{m_L+m_U} q^1(+1|x_i)} \\
&= \frac{3 + 2 \times 0.75}{4 + 2 \times 0.75 + 2 \times 0.25} \\
&= \frac{3}{4}
\end{aligned}$$

Similarly,

$$\begin{aligned}
\hat{\theta}^0_{2|+1} &= \frac{2 + 2 \times 0.75}{4 + 2 \times 0.75 + 2 \times 0.25} \\
&= \frac{7}{12} \\
\hat{\theta}^0_{1|-1} &= \frac{1 + 2 \times 0.25}{4 + 2 \times 0.25 + 2 \times 0.75} \\
&= \frac{1}{4} \\
\hat{\theta}^0_{2|-1} &= \frac{2 + 2 \times 0.25}{4 + 2 \times 0.25 + 2 \times 0.75} \\
&= \frac{5}{12}
\end{aligned}$$

4. Log-Likelihood

$$\begin{aligned}
\ln p(S;\hat{\theta}^t) = {} & 2\ln(\hat{\theta}^t_{+1}\hat{\theta}^t_{1|+1}\hat{\theta}^t_{2|+1}) + \ln(\hat{\theta}^t_{+1}\hat{\theta}^t_{1|+1}(1-\hat{\theta}^t_{2|+1})) + \ln(\hat{\theta}^t_{+1}(1-\hat{\theta}^t_{1|+1})(1-\hat{\theta}^t_{2|+1})) \\
& + \ln(\hat{\theta}^t_{-1}\hat{\theta}^t_{1|-1}(1-\hat{\theta}^t_{2|-1})) + \ln(\hat{\theta}^t_{-1}(1-\hat{\theta}^t_{1|-1})\hat{\theta}^t_{2|-1}) + 2\ln(\hat{\theta}^t_{-1}(1-\hat{\theta}^t_{1|-1})(1-\hat{\theta}^t_{2|-1})) \\
& + 2\ln(\hat{\theta}^t_{+1}\hat{\theta}^t_{1|+1}\hat{\theta}^t_{2|+1} + \hat{\theta}^t_{-1}\hat{\theta}^t_{1|-1}\hat{\theta}^t_{2|-1}) \\
& + 2\ln(\hat{\theta}^t_{+1}(1-\hat{\theta}^t_{1|+1})(1-\hat{\theta}^t_{2|+1}) + \hat{\theta}^t_{-1}(1-\hat{\theta}^t_{2|-1})(1-\hat{\theta}^t_{2|-1}))
\end{aligned}$$