Proofs of the Theorems

Wei Wang and Zhi-Hua Zhou  
National Key Laboratory for Novel Software Technology  
Nanjing University, Nanjing 210093, China  
{wangw, zhouzh}@lambda.nju.edu.cn

Proof of Theorem 1. Let $Q_i = S_1^i \oplus S_2^i$. First we prove that if each view $X_v$ ($v = 1, 2$) satisfies Tsybakov noise condition, i.e., $P_{x_v \in X_v}(|\varphi_v(x_v) - 1/2| \leq t) \leq C_3t^{1/2}$ for some finite $C_3 > 0$, $\lambda_3 > 0$ and all $0 < t \leq 1/2$, Tsybakov noise condition can also be met in $Q_i$, i.e., $P_{x_v \in Q_i}(|\varphi_v(x_v) - 1/2| \leq t) \leq C_4t^{1/2}$ for some finite $C_4 > 0$, $\lambda_4 > 0$ and all $0 < t \leq 1/2$. Suppose Tsybakov noise condition cannot be met in $Q_i$, then for $C_* = \frac{C_3}{Pr(Q_i)}$ and $\lambda_* = \lambda_3$, there exists some $0 < t_* \leq 1/2$ to satisfy that $\frac{P_{x_v \in Q_i}(|\varphi_v(x_v) - 1/2| \leq t)}{Pr(Q_i)} > C_* t_*^{1/2}$. So we get

$$Pr_{x_v \in X_v}(|\varphi_v(x_v) - 1/2| \leq t) \geq Pr_{x_v \in Q_i}(|\varphi_v(x_v) - 1/2| \leq t) > C_* t_*^{1/2}.$$  

It is in contradiction with that $X_v$ satisfies Tsybakov noise condition. Thus, we get that Tsybakov noise condition can also be met in $Q_i$. Without loss of generality, suppose that Tsybakov noise condition in all $Q_i$ and $X_v$ can be met for the same finite $C_0$ and $\lambda$.

Since $m_0 = \frac{256\lambda C}{C^2} \left(V + \log\left(\frac{16(s+1)}{\delta}\right)\right)$, according to Lemma 1 we know that $d(S_0^i, S^i) \leq C_1^i$ with probability at least $1 - \frac{\delta}{m_0(s+1)}$. With $d(S_v, S_0^i) \geq C_1d^i_\Delta(S_v, S_0^i)$, we get $d^i_\Delta(S_v^0, S^i) \leq \frac{1}{16}$. It is easy to find that $d^\Delta(S_0^i \cap S_2^i, S^i) \leq d^\Delta(S_0^i, S^i) + d^\Delta(S_2^i, S^i) \leq 1/8$ holds with probability at least $1 - \frac{\delta}{8(s+1)}$.

For $i \geq 0$, $m_{i+1}$ number of labels are queried randomly from $Q_i$. Thus, similarly according to Lemma 1 we have $d^\Delta(S_1^i \cap S_2^i \cap Q_i^i, S^i \cap Q_i) \leq 1/8$ with probability at least $1 - \frac{3}{8(s+1)}$. Let $T^i_{v+1} = S_{v+1}^i \cap Q_i$ and $\tau_{i+1} = \frac{Pr(T^i_{v+1} \oplus T^i_{v+1} \oplus S^i)}{Pr(T^i_{v+1} \oplus T^i_{v+1})} - \frac{1}{2}$, it is easy to get

$$Pr(S^i \cap (S_1^i \oplus S_2^i \oplus S^i)) - Pr(S^i \cap (S_1^i \oplus S_2^i)) = -2\tau_{i+1}Pr(S^i \cap S_2^i \cap Q_i^i).$$

Considering the non-degradation condition and $d^\Delta(S_1^i \cap S_2^i \cap Q_i^i, S^i \cap Q_i^i) = d^\Delta(S_1^i \cap Q_i^i, S^i \cap Q_i^i)$, we calculate that

$$d^\Delta(S_1^i \cap S_2^i \cap Q_i^i, S^i \cap Q_i^i) = \frac{1}{2} \left( d^\Delta(S_1^i \cap Q_i^i, S^i \cap Q_i^i) + d^\Delta(S_2^i \cap Q_i^i, S^i \cap Q_i^i) \right) + \frac{1}{2} Pr\left( S^i \cap (S_1^i \oplus S_2^i) \cap Q_i^i \right) - \frac{1}{2} Pr\left( S^i \cap (S_1^i \oplus S_2^i) \cap Q_i^i \right) - \tau_{i+1}Pr(S^i \cap S_2^i \cap Q_i^i)$$

$$= d^\Delta(S_1^i \cap S_2^i \cap Q_i^i, S^i \cap Q_i^i) - \tau_{i+1}Pr(S^i \cap S_2^i \cap Q_i^i).$$

So we have

$$d^\Delta(S_1^i \cap S_2^i \cap Q_i^i, S^i) = d^\Delta(S_1^i \cap S_2^i \cap Q_i^i, S^i \cap Q_i)Pr(Q_i^i) + d^\Delta(S_1^i \cap S_2^i \cap Q_i^i, S^i \cap Q_i)Pr(Q_i^i) - \tau_{i+1}Pr(S^i \cap S_2^i \cap Q_i^i).$$

Finally, we have

$$d^\Delta(S_1^i \cap S_2^i \cap Q_i^i, S^i) \leq \frac{1}{8} Pr(Q_i^i) + \frac{1}{8} Pr(Q_i^i) + \frac{1}{8} Pr(Q_i^i) - \tau_{i+1}Pr(S^i \cap S_2^i \cap Q_i^i).$$

$$= \frac{1}{8} Pr(Q_i^i) + \frac{1}{8} Pr(Q_i^i) + \frac{1}{8} Pr(Q_i^i) - \tau_{i+1}Pr(S^i \cap S_2^i \cap Q_i^i).$$
Considering $d_\Delta(S^*_1 \cap S^*_2 | Q_i, S^* | Q_i) P_{r}(Q_i) = P_{r}(S^*_1 \cap S^*_2 - S^*) + P_{r}(\overline{S^*_1} \cap \overline{S^*_2} - S^*)$, we have
\[ d_\Delta(S^{i+1}_1 \cap S^{i+1}_2, S^*) \leq P_{r}(S^*_1 \cap S^*_2 - S^*) + P_{r}(\overline{S^*_1} \cap \overline{S^*_2} - S^*) + \frac{1}{8} P_{r}(S^*_1 \cap S^*_2) - \tau_{i+1} P_{r}((S^{i+1}_1 \cap S^{i+1}_2) \cap Q_i). \]

Similarly, we get
\[ d_\Delta(S^{i+1}_1 \cup S^{i+1}_2, S^*) \leq P_{r}(S^*_1 \cap S^*_2 - S^*) + P_{r}(\overline{S^*_1} \cap \overline{S^*_2} - S^*) + \frac{1}{8} P_{r}(S^*_1 \cap S^*_2) + \tau_{i+1} P_{r}((S^{i+1}_1 \cap S^{i+1}_2) \cap Q_i). \]

Let $\gamma_i = \frac{P_{r}(S^*_1 \cap S^*_2) - P_{r}(\overline{S^*_1} \cap \overline{S^*_2})}{P_{r}(S^*_1 \cap S^*_2)} - \frac{1}{2}$, we have
\[ d_\Delta(S^*_1 \cap S^*_2, S^*) = \frac{d_\Delta(S^*_1 \cap S^*_2 | Q_i, S^* | Q_i) P_{r}(Q_i)}{d_\Delta(S^*_1 \cap S^*_2, S^*)} = \frac{1}{2} \gamma_i + \frac{1}{\alpha} P_{r}(S^*_1 \cap S^*_2) \leq \frac{5\alpha + 8}{8\alpha + 8}; \]
and $d_\Delta(S^*_1 \cup S^*_2, S^*) = \frac{1}{2} \gamma_i + \frac{1}{\alpha} P_{r}(S^*_1 \cap S^*_2) + \frac{1}{\alpha} P_{r}(S^*_1 \cap S^*_2) \leq \frac{5\alpha + 8}{8\alpha + 8}$.

Case 1: If $|\tau_{i+1}| \leq \gamma_i$, with respect to Definition 1, we have
\[ \frac{d_\Delta(S^{i+1}_1 \cap S^{i+1}_2, S^*)}{d_\Delta(S^*_1 \cap S^*_2, S^*)} \leq \frac{\frac{1}{8} P_{r}(S^*_1 \cap S^*_2) + |\tau_{i+1}| P_{r}(S^{i+1}_1 \cap S^{i+1}_2) + \frac{1}{\alpha} P_{r}(S^*_1 \cap S^*_2)}{\frac{1}{2} \gamma_i + \frac{1}{\alpha} P_{r}(S^*_1 \cap S^*_2)} \leq \frac{5\alpha + 8}{8\alpha + 8}; \]
\[ \frac{d_\Delta(S^{i+1}_1 \cup S^{i+1}_2, S^*)}{d_\Delta(S^*_1 \cup S^*_2, S^*)} \leq \frac{\frac{1}{8} P_{r}(S^*_1 \cap S^*_2) + |\tau_{i+1}| P_{r}(S^{i+1}_1 \cap S^{i+1}_2) + \frac{1}{\alpha} P_{r}(S^*_1 \cap S^*_2)}{\frac{1}{2} \gamma_i + \frac{1}{\alpha} P_{r}(S^*_1 \cap S^*_2)} \leq \frac{5\alpha + 8}{8\alpha + 8}; \]
\[ \frac{d_\Delta(S^{i+1}_1 \cap S^{i+1}_2, S^*)}{d_\Delta(S^*_1 \cap S^*_2, S^*)} \leq \frac{\frac{1}{8} P_{r}(S^*_1 \cap S^*_2) + \frac{1}{\alpha} P_{r}(S^*_1 \cap S^*_2)}{\frac{1}{2} \gamma_i + \frac{1}{\alpha} P_{r}(S^*_1 \cap S^*_2)} \leq \frac{\alpha + 8}{2\alpha + 8}; \]
\[ \frac{d_\Delta(S^{i+1}_1 \cup S^{i+1}_2, S^*)}{d_\Delta(S^*_1 \cup S^*_2, S^*)} \leq \frac{\frac{1}{8} P_{r}(S^*_1 \cap S^*_2) + \frac{1}{\alpha} P_{r}(S^*_1 \cap S^*_2)}{\frac{1}{2} \gamma_i + \frac{1}{\alpha} P_{r}(S^*_1 \cap S^*_2)} \leq \frac{\alpha + 8}{2\alpha + 8}; \]
\[ \frac{d_\Delta(S^{i+1}_1 \cap S^{i+1}_2, S^*)}{d_\Delta(S^*_1 \cap S^*_2, S^*)} \leq \frac{\frac{1}{8} P_{r}(S^*_1 \cap S^*_2) + \frac{1}{\alpha} P_{r}(S^*_1 \cap S^*_2)}{\frac{1}{2} \gamma_i + \frac{1}{\alpha} P_{r}(S^*_1 \cap S^*_2)} \leq \frac{\alpha + 8}{2\alpha + 8}; \]
\[ \frac{d_\Delta(S^{i+1}_1 \cup S^{i+1}_2, S^*)}{d_\Delta(S^*_1 \cup S^*_2, S^*)} \leq \frac{\frac{1}{8} P_{r}(S^*_1 \cap S^*_2) + \frac{1}{\alpha} P_{r}(S^*_1 \cap S^*_2)}{\frac{1}{2} \gamma_i + \frac{1}{\alpha} P_{r}(S^*_1 \cap S^*_2)} \leq \frac{\alpha + 8}{2\alpha + 8}; \]
Case 6: If \( \tau_{i+1} < \gamma_i \) and \(-\frac{1}{2} \leq \gamma_i < -\frac{1}{4} \), with respect to Definition 1, we have
\[
\frac{d_\Delta(S_1^{i+1} \cap S_2^{i+1}, S^*)}{d_\Delta(S_1^i \cap S_2^i, S^*)} \leq \frac{\frac{1}{\beta} \Pr(S_1^i \oplus S_2^i) + |\tau_{i+1}| \Pr(S_1^{i+1} \oplus S_2^{i+1}) + \frac{1}{2} \Pr(S_1^i \oplus S_2^i)}{\left(\frac{1}{2} + |\gamma_i|\right) \Pr(S_1^i \oplus S_2^i) + \frac{1}{\alpha} \Pr(S_1^i \oplus S_2^i)} \leq \frac{5\alpha + 8}{6\alpha + 8};
\]
Case 7: If \( \tau_{i+1} \leq -\gamma_i \) and \( 0 \leq \gamma_i \leq \frac{1}{2} \), with respect to Definition 1, we have
\[
\frac{d_\Delta(S_1^{i+1} \cap S_2^{i+1}, S^*)}{d_\Delta(S_1^i \cap S_2^i, S^*)} \leq \frac{\frac{1}{\beta} \Pr(S_1^i \oplus S_2^i) + \frac{1}{\alpha} \Pr(S_1^i \oplus S_2^i)}{\left(\frac{1}{2} + |\gamma_i|\right) \Pr(S_1^i \oplus S_2^i) + \frac{1}{\alpha} \Pr(S_1^i \oplus S_2^i)} \leq \frac{\alpha + 8}{4\alpha + 8};
\]
Case 8: If \( \tau_{i+1} > -\gamma_i \) and \(-\frac{1}{2} \leq \gamma_i \leq 0 \), with respect to Definition 1, we have
\[
\frac{d_\Delta(S_1^{i+1} \cap S_2^{i+1}, S^*)}{d_\Delta(S_1^i \cap S_2^i, S^*)} \leq \frac{\frac{1}{\beta} \Pr(S_1^i \oplus S_2^i) + \frac{1}{\alpha} \Pr(S_1^i \oplus S_2^i)}{\left(\frac{1}{2} + |\gamma_i|\right) \Pr(S_1^i \oplus S_2^i) + \frac{1}{\alpha} \Pr(S_1^i \oplus S_2^i)} \leq \frac{\alpha + 8}{4\alpha + 8};
\]
Thus, after the \((i + 1)\)-th round, either \( \frac{d_\Delta(S_1^i \cap S_2^i, S^*)}{d_\Delta(S_1^i \cap S_2^i, S^*)} \leq \frac{5\alpha + 8}{6\alpha + 8} \) or \( \frac{d_\Delta(S_1^{i+1} \cap S_2^{i+1}, S^*)}{d_\Delta(S_1^i \cap S_2^i, S^*)} \leq \frac{5\alpha + 8}{6\alpha + 8} \)
holds. Hence, we have \( d_\Delta(S_1^i \cap S_2^i, S^*) \leq \frac{1}{8} \left(\frac{5\alpha + 8}{6\alpha + 8}\right)^{i/2} \) or \( d_\Delta(S_1^{i+1} \cap S_2^{i+1}, S^*) \leq \frac{1}{8} \left(\frac{5\alpha + 8}{6\alpha + 8}\right)^{i/2} \)
with probability at least \( 1 - \delta \). When \( s = \lceil \frac{2\log C_2}{\beta} \rceil \), where \( C_2 = \frac{5\alpha + 8}{6\alpha + 8} \) is a constant less than 1, we have either \( d_\Delta(S_1^i \cap S_2^i, S^*) \leq \epsilon \) or \( d_\Delta(S_1^{i+1} \cap S_2^{i+1}, S^*) \leq \epsilon \) with probability at least \( 1 - \delta \). Thus, considering \( R(h_1^+) - R(S^*) = R(S_1^i \cap S_2^i) - R(S^*) \leq d_\Delta(S_1^i \cap S_2^i, S^*) \) and \( R(h_1^-) - R(S^*) = R(S_1^i \cap S_2^i) - R(S^*) \leq d_\Delta(S_1^{i+1} \cap S_2^{i+1}, S^*) \), we have either \( R(h_1^+) \leq R(S^*) + \epsilon \) or \( R(h_1^-) \leq R(S^*) + \epsilon \).

Proof of Lemma 2. We apply \( S_1^i \) and \( S_2^i \) to the unlabeled instances set and identify the contention point set. Then we query for labels of \( \frac{2\log(\frac{4}{\beta})}{\beta^2} \) instances drawn randomly from the contention points set. With these labels we estimate the empirical value \( \hat{P}_1 \) of \( \frac{\Pr(x : x \in S_1^i \cap S_2^i \land y(x) = 1)}{\Pr(S_1^i \cap S_2^i)} \) and the empirical value \( \hat{P}_2 \) of \( \frac{\Pr(x : x \in S_1^i \cap S_2^i \land y(x) = 0)}{\Pr(S_1^i \cap S_2^i)} \). By Chernoff bound, with number of \( \frac{2\log(\frac{4}{\beta})}{\beta^2} \) labels we have the following two equations with probability at least \( 1 - \delta \).
\[
\hat{P}_1 \in \left[ \frac{\Pr(\{x : x \in S_1^i \cap S_2^i \land y(x) = 1\})}{\Pr(S_1^i \cap S_2^i)} - \frac{\beta}{2} \frac{\Pr(\{x : x \in S_1^i \cap S_2^i \land y(x) = 1\})}{\Pr(S_1^i \cap S_2^i)} + \frac{\beta}{2} \right];
\hat{P}_2 \in \left[ \frac{\Pr(\{x : x \in S_1^i \cap S_2^i \land y(x) = 0\})}{\Pr(S_1^i \cap S_2^i)} - \frac{\beta}{2} \frac{\Pr(\{x : x \in S_1^i \cap S_2^i \land y(x) = 0\})}{\Pr(S_1^i \cap S_2^i)} + \frac{\beta}{2} \right];
\]
If \( \hat{P}_1 \leq \hat{P}_2 \), we get \( \Pr(\{x : x \in S_1^i \cap S_2^i \land y(x) = 1\}) \leq \Pr(\{x : x \in S_1^i \cap S_2^i \land y(x) = 0\}) \) with probability at least \( 1 - \delta \); otherwise, we get \( \Pr(\{x : x \in S_1^i \cap S_2^i \land y(x) = 1\}) > \Pr(\{x : x \in S_1^i \cap S_2^i \land y(x) = 0\}) \) with probability at least \( 1 - \delta \).

Proof of Theorem 2. According to Theorem 1, by requesting \( O(\log \frac{1}{\delta}) \) labels the multi-view active learning in Table 1 can get either \( R(h_1^+) \leq R(S^*) + \epsilon \) or \( R(h_1^-) \leq R(S^*) + \epsilon \) with probability at least \( 1 - \frac{\delta}{2} \). According to Lemma 2, by requesting \( \frac{2\log(\frac{4}{\delta})}{\beta^2} \) labels we can decide correctly whether \( \Pr(\{x : x \in S_1^i \cap S_2^i \land y(x) = 1\}) \) or \( \Pr(\{x : x \in S_1^i \cap S_2^i \land y(x) = 0\}) \) is smaller with probability at least \( 1 - \frac{\delta}{2} \).

Case 1: If \( \Pr(\{x : x \in S_1^i \cap S_2^i \land y(x) = 1\}) \leq \Pr(\{x : x \in S_1^i \cap S_2^i \land y(x) = 0\}) \), we have \( R(h_1^-) \leq R(h_1^+) \). Thus, we get \( R(h_1--) \leq R(S^*) + \epsilon \) with probability at least \( 1 - \delta \).
Case 2: If \( Pr(\{x : x \in S_i^c + S_j^c \land y(x) = 1\}) > Pr(\{x : x \in S_i^c + S_j^c \land y(x) = 0\}) \), we have \( R(h^*_i) < R(h^*_j) \). Thus, we get \( R(h^*_i) \leq R(S^*) + \epsilon \) with probability at least \( 1 - \delta \).

The total number of labels to be requested is \( \tilde{O}(\log \frac{1}{\epsilon}) + \frac{2\log(\frac{\delta}{\epsilon})}{\delta} = \tilde{O}(\log \frac{1}{\epsilon}) \). □

Proof of Theorem 3. Since \( Pr(S_i^c + S_j^c) \leq 1 \), with the following equation

\[
\left| \frac{Pr(\{x : x \in S_i^c + S_j^c \land y(x) = 1\})}{Pr(S_i^c + S_j^c)} - \frac{Pr(\{x : x \in S_i^c + S_j^c \land y(x) = 0\})}{Pr(S_i^c + S_j^c)} \right| = O(\epsilon)
\]

we have \[ Pr(\{x : x \in S_i^c + S_j^c \land y(x) = 1\}) - Pr(\{x : x \in S_i^c + S_j^c \land y(x) = 0\}) \] = \( O(\epsilon) \). So it is easy to get \( |R(h^*_i) - R(h^*_j)| = O(\epsilon) \). According to Theorem 1, by requesting \( \tilde{O}(\log \frac{1}{\epsilon}) \) labels we can get either \( R(h^*_i) \leq R(S^*) + \epsilon \) or \( R(h^*_j) \leq R(S^*) + \epsilon \) with probability at least \( 1 - \delta \). Thus, we get that \( h^*_i \) and \( h^*_j \) satisfy either (a) or (b) with probability at least \( 1 - \delta \). □

Proof of Theorem 5. According to Theorem 4, by requesting \( \tilde{O}(\log \frac{1}{\epsilon}) \) labels the multi-view active learning in Table 1 can get either \( R(h^*_i) \leq R(S_i^c \cap S_j^c) + \epsilon \) or \( R(h^*_j) \leq R(S_i^c \cap S_j^c) + \epsilon \) with probability at least \( 1 - \delta \). According to Lemma 2, by requesting \( \frac{2\log(\frac{\delta}{\epsilon})}{\delta} \) labels we can decide correctly whether \( Pr(\{x : x \in S_i^c + S_j^c \land y(x) = 1\}) \) or \( Pr(\{x : x \in S_i^c + S_j^c \land y(x) = 0\}) \) is smaller with probability at least \( 1 - \delta \).

Case 1: If \( Pr(\{x : x \in S_i^c + S_j^c \land y(x) = 1\}) \leq Pr(\{x : x \in S_i^c + S_j^c \land y(x) = 0\}) \), we have \( R(h^*_i) \leq R(h^*_j) \). Thus, we get \( R(h^*_i) \leq R(S_i^c \cap S_j^c) + \epsilon \) with probability at least \( 1 - \delta \).

Case 2: If \( Pr(\{x : x \in S_i^c + S_j^c \land y(x) = 1\}) > Pr(\{x : x \in S_i^c + S_j^c \land y(x) = 0\}) \), we have \( R(h^*_i) < R(h^*_j) \). Thus, we get \( R(h^*_i) \leq R(S_i^c \cap S_j^c) + \epsilon \) with probability at least \( 1 - \delta \).

The total number of labels to be requested is \( \tilde{O}(\log \frac{1}{\epsilon}) + \frac{2\log(\frac{\delta}{\epsilon})}{\delta} = \tilde{O}(\log \frac{1}{\epsilon}) \). □

Proof of Corollary 1. According to Theorem 5 we know that by requesting \( \tilde{O}(\log \frac{1}{\epsilon}) \) labels the multi-view active learning in Table 1 will generate a classifier whose error rate is no larger than \( R(S_i^c \cap S_j^c) + \frac{\epsilon}{2} \) with probability at least \( 1 - \delta \). Considering that

\[
R(S_i^c \cap S_j^c) - R(S^*) = \int_{(S_i^c \cap S_j^c) \Delta S^*} [2\phi_S(x,v) - 1]p_x d_x \leq Pr(S_i^c + S_j^c),
\]

we have \( R(S_i^c \cap S_j^c) \leq R(S^*) + \frac{\epsilon}{2} \). Thus, we get that \( R(S_i^c \cap S_j^c) + \frac{\epsilon}{2} \) is no larger than \( R(S^*) + \epsilon \). □

Proof of Theorem 6. After the \( i \)-th round in Table 2, the number of training examples in \( \mathcal{L} \) is \( \sum_{b=0}^{\beta} 2^b m_b = (2^{i+1} - 1)m_i \). While in the \( (i+1) \)-th round, we randomly query \( (2^{i+1} - 1)m_i \) labels from the region of \( \mathcal{Q}_i \) and add them into \( \mathcal{L} \). So in the \( (i+1) \)-th round, the number of training examples for \( S_i^{l+1} \) \( (v = 1, 2) \) drawn randomly from region of \( \mathcal{Q}_i \) is larger than the number of whole training examples for \( S_i^c \). Since the optimal Bayes classifier \( c_v \) belongs to \( \mathcal{H}_v \), according to the standard PAC-model, it is easy to know that \( d(S_v^{l+1}|\mathcal{Q}_i, S^*|\mathcal{Q}_i) \leq d(S_v^c|\mathcal{Q}_i, S^*|\mathcal{Q}_i) \) can be met for any \( \varphi_v \), where \( d(S_v|\mathcal{Q}_i, S^*|\mathcal{Q}_i) \) is defined as

\[
d(S_v|\mathcal{Q}_i, S^*|\mathcal{Q}_i) = d(S_v|\mathcal{Q}_i) - R(S^*|\mathcal{Q}_i) = \int_{(S_v \cap \mathcal{Q}_i) \Delta (S^* \cap \mathcal{Q}_i)} [2\phi_S(x,v) - 1]p_x d_x / Pr(\mathcal{Q}_i).
\]

So, by setting \( \varphi_v \in \{0, 1\} \), we get \( d(S_v^{l+1}|\mathcal{Q}_i, S^*|\mathcal{Q}_i) \leq d(S_v^c|\mathcal{Q}_i, S^*|\mathcal{Q}_i) \), which implies the non-degradation condition. Thus, with the proof of Theorem 1, we get Theorem 6 proved. □

Proof of Theorem 9. Similarly to the proof of Theorem 4 and Theorem 6, we know that by requesting \( \tilde{O}(\frac{1}{\epsilon}) \) labels the multi-view active learning in Table 2 can get either \( R(h^*_i) \leq R(S_i^c \cap S_j^c) + \epsilon \) or \( R(h^*_j) \leq R(S_i^c \cap S_j^c) + \epsilon \) with probability at least \( 1 - \delta \). According to Lemma 2, by requesting
\[ \frac{2 \log(\frac{\delta}{\beta^2})}{\beta^2} \] labels we can decide correctly whether \( R(h^*_+) \) or \( R(h^-) \) is smaller with probability at least \( 1 - \frac{\delta}{2} \). Thus, we can get a classifiers whose error rate is no larger than \( R(S^*_1 \cap S^*_2) + \varepsilon \) with probability at least \( 1 - \delta \). The total number of labels to be requested is \( \tilde{O}(\frac{1}{\varepsilon}) + \frac{2 \log(\frac{\delta}{\beta^2})}{\beta^2} = \tilde{O}(\frac{1}{\varepsilon}) \). \[ \square \]

**Proof of Corollary 2.** According to Theorem 9 we know that by requesting \( \tilde{O}(\frac{1}{\varepsilon}) \) labels the multi-view active learning in Table 2 will generate a classifier whose error rate is no larger than \( R(S^*_1 \cap S^*_2) + \varepsilon \) with probability at least \( 1 - \delta \). With the proof of Corollary 1, we get that \( R(S^*_1 \cap S^*_2) + \varepsilon \) is no larger than \( R(S^*_v) + \varepsilon \). \[ \square \]