A. Proof of Lemma 1

From the assumption (2), $\mathcal{D}_{\mathrm{S}} = \{(\boldsymbol{x}_{\mathrm{S},i}, \boldsymbol{x}'_{\mathrm{S},i})\}_{i=1}^{n_{\mathrm{S}}} \sim p_{\mathrm{S}}(\boldsymbol{x}, \boldsymbol{x}')$. In order to decompose pairwise data into pointwise, marginalize $p_{\mathrm{S}}(\boldsymbol{x}, \boldsymbol{x}')$ with respect to $\boldsymbol{x}'$:

$$\begin{aligned}
\int p_{\mathrm{S}}(\boldsymbol{x}, \boldsymbol{x}') d\boldsymbol{x}' &= \frac{\pi_+^2}{\pi_+^2 + \pi_-^2} p_+(\boldsymbol{x}) \int p_+(\boldsymbol{x}') d\boldsymbol{x}' + \frac{\pi_-^2}{\pi_+^2 + \pi_-^2} p_-(\boldsymbol{x}) \int p_-(\boldsymbol{x}') d\boldsymbol{x}' \\
&= \frac{\pi_+^2}{\pi_+^2 + \pi_-^2} p_+(\boldsymbol{x}) \int \frac{p(\boldsymbol{x}', y = +1)}{p(y = +1)} d\boldsymbol{x}' + \frac{\pi_-^2}{\pi_+^2 + \pi_-^2} p_-(\boldsymbol{x}) \int \frac{p(\boldsymbol{x}', y = -1)}{p(y = -1)} d\boldsymbol{x}' \\
&= \frac{\pi_+^2}{\pi_+^2 + \pi_-^2} p_+(\boldsymbol{x}) \frac{p(y = +1)}{p(y = +1)} + \frac{\pi_-^2}{\pi_+^2 + \pi_-^2} p_-(\boldsymbol{x}) \frac{p(y = -1)}{p(y = -1)} \\
&= \frac{\pi_+^2}{\pi_+^2 + \pi_-^2} p_+(\boldsymbol{x}) + \frac{\pi_-^2}{\pi_+^2 + \pi_-^2} p_-(\boldsymbol{x}) \\
&= \widetilde{p}_{\mathrm{S}}(\boldsymbol{x}).
\end{aligned}$$

Since a pair $(\boldsymbol{x}_{\mathrm{S},i}, \boldsymbol{x}'_{\mathrm{S},i}) \in \mathcal{D}_{\mathrm{S}}$ is independently and identically drawn, both $\boldsymbol{x}_{\mathrm{S},i}$ and $\boldsymbol{x}'_{\mathrm{S},i}$ are drawn following $\widetilde{p}_{\mathrm{S}}$. □

B. Proof of Theorem 1

To prove Theorem 1, it is convenient to begin with the following Lemma 3.

Lemma 3. *The classification risk (1) can be equivalently expressed as*

$$\begin{aligned}
R_{\mathrm{PSD},\ell}(f) = &\frac{\pi_+}{2\pi_-} \mathop{\mathbb{E}}_{X \sim p_+} \left[\widetilde{\ell}(f(X)) \right] \\
&+ \pi_{\mathrm{S}} \mathop{\mathbb{E}}_{(X,X') \sim p_{\mathrm{S}}} \left[-\frac{\pi_+}{2\pi_-} \frac{\ell(f(X), +1) + \ell(f(X'), +1)}{2} + \frac{1 + \pi_-}{2\pi_-} \frac{\ell(f(X), -1) + \ell(f(X'), -1)}{2} \right] \\
&+ \pi_{\mathrm{D}} \mathop{\mathbb{E}}_{(X,X') \sim p_{\mathrm{D}}} \left[\frac{\ell(f(X), -1) + \ell(f(X'), +1)}{2} \right],
\end{aligned} \tag{B.1}$$

where $\mathbb{E}_{X \sim p_+}[\cdot]$, $\mathbb{E}_{(X,X') \sim p_{\mathrm{S}}}[\cdot]$, and $\mathbb{E}_{(X,X') \sim p_{\mathrm{D}}}[\cdot]$ denote expectations over $p_+(X)$, $p_{\mathrm{S}}(X, X')$, and $p_{\mathrm{D}}(X, X')$, respectively.

Note that the definitions of p_{D} and π_{D} are first given in Eq. (12).

Proof. Eq. (1) can be transformed into pairwise fashion:

$$\begin{aligned}
\mathop{\mathbb{E}}_{(X,Y) \sim p} [\ell(f(X), Y)] &= \mathop{\mathbb{E}}_{(X,Y) \sim p} \left[\frac{\ell(f(X), Y)}{2} \right] + \mathop{\mathbb{E}}_{(X',Y') \sim p} \left[\frac{\ell(f(X'), Y')}{2} \right] \\
&= \mathop{\mathbb{E}}_{(X,Y),(X',Y') \sim p} \left[\frac{\ell(f(X), Y) + \ell(f(X'), Y')}{2} \right].
\end{aligned} \tag{B.2}$$

Both pairs (X, Y) and (X', Y') are independently and identically distributed from the joint distribution $p(\boldsymbol{x}, y)$. Thus, Eq. (B.2) can be further decomposed:

$$\begin{aligned}
&\mathop{\mathbb{E}}_{(X,Y),(X',Y') \sim p} \left[\frac{\ell(f(X), Y) + \ell(f(X'), Y')}{2} \right] \\
&= \sum_{y,y'} \int \frac{\ell(f(\boldsymbol{x}), y) + \ell(f(\boldsymbol{x}'), y')}{2} p(\boldsymbol{x}, y) p(\boldsymbol{x}', y') d\boldsymbol{x} d\boldsymbol{x}' \\
&= \pi_+^2 \int \frac{\ell(f(\boldsymbol{x}), +1) + \ell(f(\boldsymbol{x}'), +1)}{2} p_+(\boldsymbol{x}) p_+(\boldsymbol{x}') d\boldsymbol{x} d\boldsymbol{x}' \\
&\quad + \pi_-^2 \int \frac{\ell(f(\boldsymbol{x}), -1) + \ell(f(\boldsymbol{x}'), -1)}{2} p_-(\boldsymbol{x}) p_-(\boldsymbol{x}') d\boldsymbol{x} d\boldsymbol{x}'
\end{aligned}$$

$$
\begin{aligned}
&+\pi_+\pi_- \int \frac{\ell(f(\boldsymbol{x}), +1) + \ell(f(\boldsymbol{x}'), -1)}{2} p_+(\boldsymbol{x}) p_-(\boldsymbol{x}') d\boldsymbol{x} d\boldsymbol{x}' \\
&+\pi_+\pi_- \int \frac{\ell(f(\boldsymbol{x}), -1) + \ell(f(\boldsymbol{x}'), +1)}{2} p_-(\boldsymbol{x}) p_+(\boldsymbol{x}') d\boldsymbol{x} d\boldsymbol{x}' \\
=\ &\pi_+^2 \mathop{\mathbb{E}}_{X, X' \sim p_+} \left[\frac{\ell(f(X), +1) + \ell(f(X'), +1)}{2} \right] \\
&+\pi_-^2 \mathop{\mathbb{E}}_{X, X' \sim p_-} \left[\frac{\ell(f(X), -1) + \ell(f(X'), -1)}{2} \right] \\
&+\pi_+\pi_- \mathop{\mathbb{E}}_{X \sim p_+, X' \sim p_-} \left[\frac{\ell(f(X), +1) + \ell(f(X'), -1)}{2} \right] \\
&+\pi_+\pi_- \mathop{\mathbb{E}}_{X \sim p_-, X' \sim p_+} \left[\frac{\ell(f(X), -1) + \ell(f(X'), +1)}{2} \right]. \qquad \text{(B.3)}
\end{aligned}
$$

Using Eq. (2), the following equation is obtained:

$$
\begin{aligned}
&\pi_{\mathrm{S}} \mathop{\mathbb{E}}_{(X, X') \sim p_{\mathrm{S}}} \left[\frac{\ell(f(X), -1) + \ell(f(X'), -1)}{2} \right] \\
&= \pi_+^2 \mathop{\mathbb{E}}_{X, X' \sim p_+} \left[\frac{\ell(f(X), -1) + \ell(f(X'), -1)}{2} \right] + \pi_-^2 \mathop{\mathbb{E}}_{X, X' \sim p_-} \left[\frac{\ell(f(X), -1) + \ell(f(X'), -1)}{2} \right]. \qquad \text{(B.4)}
\end{aligned}
$$

Similarly, the following equation is obtained from Eq. (12):

$$
\begin{aligned}
&2\pi_-\pi_+ \mathop{\mathbb{E}}_{(X, X') \sim p_{\mathrm{D}}} \left[\frac{\ell(f(X), -1) + \ell(f(X'), +1)}{2} \right] \\
&= \pi_+\pi_- \mathop{\mathbb{E}}_{X \sim p_+, X' \sim p_-} \left[\frac{\ell(f(X), -1) + \ell(f(X'), +1)}{2} \right] + \pi_+\pi_- \mathop{\mathbb{E}}_{X \sim p_-, X \sim p_+} \left[\frac{\ell(f(X), -1) + \ell(f(X'), +1)}{2} \right]. \qquad \text{(B.5)}
\end{aligned}
$$

Combining Eqs. (B.3), (B.4) and (B.5), the expected risk $R(f)$ is written as

$$
\begin{aligned}
R(f) &= \pi_{\mathrm{S}} \mathop{\mathbb{E}}_{(X, X') \sim p_{\mathrm{S}}} \left[\frac{\ell(f(X), -1) + \ell(f(X'), -1)}{2} \right] + \pi_{\mathrm{D}} \mathop{\mathbb{E}}_{(X, X') \sim p_{\mathrm{D}}} \left[\frac{\ell(f(X), -1) + \ell(f(X'), +1)}{2} \right] \\
&+ \pi_+^2 \mathop{\mathbb{E}}_{X, X' \sim p_+} \left[\frac{\widetilde{\ell}(f(X)) + \widetilde{\ell}(f(X'))}{2} \right] + \pi_+\pi_- \mathop{\mathbb{E}}_{X \sim p_+, X' \sim p_-} \left[\frac{\widetilde{\ell}(f(X)) - \widetilde{\ell}(f(X'))}{2} \right]. \qquad \text{(B.6)}
\end{aligned}
$$

Here

- the second term on the RHS of Eq. (B.4) is substituted into the second term in the last line of Eq. (B.3).
- the second term on the RHS of Eq. (B.5) is substituted into the fourth term in the last line of Eq. (B.3).

On the third and fourth term on the RHS of Eq. (B.6),

$$
\begin{aligned}
&\pi_+^2 \mathop{\mathbb{E}}_{X, X' \sim p_+} \left[\frac{\widetilde{\ell}(f(X)) + \widetilde{\ell}(f(X'))}{2} \right] + \pi_+\pi_- \mathop{\mathbb{E}}_{X \sim p_+, X' \sim p_-} \left[\frac{\widetilde{\ell}(f(X)) - \widetilde{\ell}(f(X'))}{2} \right] \\
&= \pi_+^2 \left\{ \frac{1}{2} \mathop{\mathbb{E}}_{X \sim p_+} \left[\widetilde{\ell}(f(X)) \right] + \frac{1}{2} \mathop{\mathbb{E}}_{X' \sim p_+} \left[\widetilde{\ell}(f(X')) \right] \right\} + \pi_+\pi_- \left\{ \frac{1}{2} \mathop{\mathbb{E}}_{X \sim p_+} \left[\widetilde{\ell}(f(X)) \right] - \frac{1}{2} \mathop{\mathbb{E}}_{X' \sim p_-} \left[\widetilde{\ell}(f(X')) \right] \right\} \\
&= \pi_+^2 \mathop{\mathbb{E}}_{X \sim p_+} \left[\widetilde{\ell}(f(X)) \right] + \frac{1}{2}\pi_+\pi_- \mathop{\mathbb{E}}_{X \sim p_+} \left[\widetilde{\ell}(f(X)) \right] - \frac{1}{2}\pi_+\pi_- \mathop{\mathbb{E}}_{X \sim p_-} \left[\widetilde{\ell}(f(X)) \right] \\
&= \frac{\pi_+(1+\pi_+)}{2} \mathop{\mathbb{E}}_{X \sim p_+} \left[\widetilde{\ell}(f(X)) \right] - \frac{\pi_+(1-\pi_+)}{2} \mathop{\mathbb{E}}_{X \sim p_-} \left[\widetilde{\ell}(f(X)) \right] \\
&= (*). \qquad \text{(B.7)}
\end{aligned}
$$

Here similarly to derivation of Eq. (B.4),

$$\begin{aligned}
&\pi_{\mathrm{S}} \mathop{\mathbb{E}}_{(X,X')\sim p_{\mathrm{S}}}\left[\frac{\widetilde{\ell}(f(X))+\widetilde{\ell}(f(X'))}{2}\right] \\
&= \pi_+^2 \mathop{\mathbb{E}}_{X,X'\sim p_+}\left[\frac{\widetilde{\ell}(f(X))+\widetilde{\ell}(f(X'))}{2}\right] + \pi_-^2 \mathop{\mathbb{E}}_{X,X'\sim p_-}\left[\frac{\widetilde{\ell}(f(X))+\widetilde{\ell}(f(X'))}{2}\right] \\
&= \pi_+^2 \mathop{\mathbb{E}}_{X\sim p_+}\left[\widetilde{\ell}(f(X))\right] + \pi_-^2 \mathop{\mathbb{E}}_{X\sim p_-}\left[\widetilde{\ell}(f(X))\right]. \qquad \text{(B.8)}
\end{aligned}$$

Combining Eqs. (B.7) and (B.8),

$$\begin{aligned}
(*) &= \frac{\pi_+(1+\pi_+)}{2} \mathop{\mathbb{E}}_{X\sim p_+}\left[\widetilde{\ell}(f(X))\right] - \frac{\pi_+(1-\pi_+)}{2\pi_-^2}\left\{-\pi_+^2 \mathop{\mathbb{E}}_{X\sim p_+}\left[\widetilde{\ell}(f(X))\right] + \pi_{\mathrm{S}} \mathop{\mathbb{E}}_{(X,X')\sim p_{\mathrm{S}}}\left[\frac{\widetilde{\ell}(f(X))+\widetilde{\ell}(f(X'))}{2}\right]\right\} \\
&= \frac{\pi_+}{2\pi_-} \mathop{\mathbb{E}}_{X\sim p_+}\left[\widetilde{\ell}(f(X))\right] - \frac{\pi_+\pi_{\mathrm{S}}}{2\pi_-} \mathop{\mathbb{E}}_{(X,X')\sim p_{\mathrm{S}}}\left[\frac{\widetilde{\ell}(f(X))+\widetilde{\ell}(f(X'))}{2}\right]. \qquad \text{(B.9)}
\end{aligned}$$

Finally from Eqs. (B.6) and (B.9), the expected risk $R(f)$ is written as

$$\begin{aligned}
R(f) &= \pi_{\mathrm{S}} \mathop{\mathbb{E}}_{(X,X')\sim p_{\mathrm{S}}}\left[\frac{\ell(f(X),-1)+\ell(f(X'),-1)}{2}\right] + \pi_{\mathrm{D}} \mathop{\mathbb{E}}_{(X,X')\sim p_{\mathrm{D}}}\left[\frac{\ell(f(X),-1)+\ell(f(X'),+1)}{2}\right] \\
&\quad + \frac{\pi_+}{2\pi_-} \mathop{\mathbb{E}}_{X\sim p_+}\left[\widetilde{\ell}(f(X))\right] - \frac{\pi_+\pi_{\mathrm{S}}}{2\pi_-} \mathop{\mathbb{E}}_{(X,X')\sim p_{\mathrm{S}}}\left[\frac{\widetilde{\ell}(f(X))+\widetilde{\ell}(f(X'))}{2}\right] \\
&= R_{\mathrm{PSD},\ell}(f). \qquad \text{(B.10)}
\end{aligned}$$

□

Now we give a proof for Theorem 1.

Proof of Theorem 1. By Lemma 3, it is enough to show $R_{\mathrm{SU},\ell}(f) = R_{\mathrm{PSD},\ell}(f)$.

From Eq. (11),

$$\begin{aligned}
&\mathop{\mathbb{E}}_{X\sim p}\left[\frac{\ell(f(X),-1)+\ell(f(X),+1)}{2}\right] \\
&= \mathop{\mathbb{E}}_{X\sim p}\left[\frac{\ell(f(X),-1)}{2}\right] + \mathop{\mathbb{E}}_{X\sim p}\left[\frac{\ell(f(X)),+1}{2}\right] \qquad (\because \text{ linearity of the expectation}) \\
&= \mathop{\mathbb{E}}_{X\sim p}\left[\frac{\ell(f(X),-1)}{2}\right] + \mathop{\mathbb{E}}_{X'\sim p}\left[\frac{\ell(f(X')),+1}{2}\right] \\
&= \mathop{\mathbb{E}}_{X,X'\sim p}\left[\frac{\ell(f(X),-1)+\ell(f(X'),+1)}{2}\right] \\
&= \pi_{\mathrm{S}} \mathop{\mathbb{E}}_{(X,X')\sim p_{\mathrm{S}}}\left[\frac{\ell(f(X),-1)+\ell(f(X'),+1)}{2}\right] + \pi_{\mathrm{D}} \mathop{\mathbb{E}}_{(X,X')\sim p_{\mathrm{D}}}\left[\frac{\ell(f(X),-1)+\ell(f(X'),+1)}{2}\right], \qquad \text{(B.11)}
\end{aligned}$$

where $\mathbb{E}_{X\sim p}[\cdot]$ denotes the expectation over the marginal distribution $p(X)$ and the last equality is obtained from Eq. (11). Eq. (B.11) produces an alternative expression of the expectation over p_{D} (the third term on the RHS of Eq. (B.1)):

$$\begin{aligned}
&\pi_{\mathrm{D}} \mathop{\mathbb{E}}_{(X,X')\sim p_{\mathrm{D}}}\left[\frac{\ell(f(X),-1)+\ell(f(X'),+1)}{2}\right] \\
&= \mathop{\mathbb{E}}_{X\sim p}\left[\frac{\ell(f(X),-1)+\ell(f(X),+1)}{2}\right] - \pi_{\mathrm{S}} \mathop{\mathbb{E}}_{(X,X')\sim p_{\mathrm{S}}}\left[\frac{\ell(f(X),-1)+\ell(f(X'),+1)}{2}\right]. \qquad \text{(B.12)}
\end{aligned}$$

Next, we obtain an alternative expression of the expectation over positive data (the first term in RHS of Eq. (B.1)). The following two equations (B.13) and (B.14) are useful:

$$\mathop{\mathbb{E}}_{X\sim p}\left[\widetilde{\ell}(f(X))\right] = \pi_+ \mathop{\mathbb{E}}_{X\sim p_+}\left[\widetilde{\ell}(f(X))\right] + \pi_- \mathop{\mathbb{E}}_{X\sim p_-}\left[\widetilde{\ell}(f(X))\right], \tag{B.13}$$

which can simply be obtained from Eq. (3).

$$\begin{aligned}
\pi_\mathrm{S} \mathop{\mathbb{E}}_{(X,X')\sim p_\mathrm{S}}\left[\frac{\widetilde{\ell}(f(X))+\widetilde{\ell}(f(X'))}{2}\right] &= \pi_+^2 \mathop{\mathbb{E}}_{X,X'\sim p_+}\left[\frac{\widetilde{\ell}(f(X))+\widetilde{\ell}(f(X'))}{2}\right] + \pi_-^2 \mathop{\mathbb{E}}_{X,X'\sim p_-}\left[\frac{\widetilde{\ell}(f(X))+\widetilde{\ell}(f(X'))}{2}\right] \\
&= \pi_+^2 \left\{\frac{1}{2}\mathop{\mathbb{E}}_{X\sim p_+}\left[\widetilde{\ell}(f(X))\right] + \frac{1}{2}\mathop{\mathbb{E}}_{X'\sim p_+}\left[\widetilde{\ell}(f(X'))\right]\right\} \\
&\quad + \pi_-^2 \left\{\frac{1}{2}\mathop{\mathbb{E}}_{X\sim p_-}\left[\widetilde{\ell}(f(X))\right] + \frac{1}{2}\mathop{\mathbb{E}}_{X'\sim p_-}\left[\widetilde{\ell}(f(X'))\right]\right\} \\
&= \pi_+^2 \mathop{\mathbb{E}}_{X\sim p_+}\left[\widetilde{\ell}(f(X))\right] + \pi_-^2 \mathop{\mathbb{E}}_{X\sim p_-}\left[\widetilde{\ell}(f(X))\right],
\end{aligned} \tag{B.14}$$

which is obtained from Eq. (2). By calculating (B.14) $-\ \pi_- \times$ (B.13) and organizing, we obtain

$$\frac{\pi_+}{2\pi_-}\mathop{\mathbb{E}}_{X\sim p_+}\left[\widetilde{\ell}(f(X))\right] = \frac{\pi_\mathrm{S}}{2\pi_-(2\pi_+-1)}\mathop{\mathbb{E}}_{(X,X')\sim p_\mathrm{S}}\left[\frac{\widetilde{\ell}(f(X))+\widetilde{\ell}(f(X'))}{2}\right] - \frac{1}{2(2\pi_+-1)}\mathop{\mathbb{E}}_{X\sim p}\left[\widetilde{\ell}(f(X))\right]. \tag{B.15}$$

Substituting Eqs. (B.12) and (B.15) into Eq. (B.1),

$$\begin{aligned}
R_{\mathrm{PSD},\ell}(f) &= \frac{\pi_\mathrm{S}}{2\pi_-(2\pi_+-1)}\mathop{\mathbb{E}}_{(X,X')\sim p_\mathrm{S}}\left[\frac{\widetilde{\ell}(f(X))+\widetilde{\ell}(f(X'))}{2}\right] - \frac{1}{2(2\pi_+-1)}\mathop{\mathbb{E}}_{X\sim p}\left[\widetilde{\ell}(f(X))\right] \\
&\quad + \pi_\mathrm{S}\mathop{\mathbb{E}}_{(X,X')\sim p_\mathrm{S}}\left[-\frac{\pi_+}{2\pi_-}\frac{\ell(f(X),+1)+\ell(f(X'),+1)}{2} + \frac{1+\pi_-}{2\pi_-}\frac{\ell(f(X),-1)+\ell(f(X'),-1)}{2}\right] \\
&\quad + \mathop{\mathbb{E}}_{X\sim p}\left[\frac{\ell(f(X),-1)+\ell(f(X),+1)}{2}\right] - \pi_\mathrm{S}\mathop{\mathbb{E}}_{(X,X')\sim p_\mathrm{S}}\left[\frac{\ell(f(X),-1)+\ell(f(X'),+1)}{2}\right] \\
&= \pi_\mathrm{S}\mathop{\mathbb{E}}_{(X,X')\sim p_\mathrm{S}}\left[\frac{1+2\pi_+}{4(2\pi_+-1)}\widetilde{\ell}(f(X)) + \frac{1+2\pi_-}{4(2\pi_+-1)}\widetilde{\ell}(f(X'))\right] \\
&\quad + \mathop{\mathbb{E}}_{X\sim p}\left[-\frac{\pi_-}{2\pi_+-1}\ell(f(X),+1) + \frac{\pi_+}{2\pi_+-1}\ell(f(X),-1)\right] \\
&= \pi_\mathrm{S}\mathop{\mathbb{E}}_{(X,X')\sim p_\mathrm{S}}\left[\frac{\frac{1}{2\pi_+-1}\widetilde{\ell}(f(X)) + \frac{1}{2\pi_+-1}\widetilde{\ell}(f(X'))}{2}\right] \\
&\quad + \mathop{\mathbb{E}}_{X\sim p}\left[-\frac{\pi_-}{2\pi_+-1}\ell(f(X),+1) + \frac{\pi_+}{2\pi_+-1}\ell(f(X),-1)\right] \\
&= R_{\mathrm{SU},\ell}(f),
\end{aligned} \tag{B.16}$$

which concludes the proof. The third equality of Eq. (B.16) holds because X and X' are symmetric and

$$\frac{1+2\pi_+}{4(2\pi_+-1)}\widetilde{\ell}(\cdot) + \frac{1+2\pi_-}{4(2\pi_+-1)}\widetilde{\ell}(\cdot) = \frac{1}{2\pi_+-1}\widetilde{\ell}(\cdot) = \frac{\frac{1}{2\pi_+-1}\widetilde{\ell}(\cdot) + \frac{1}{2\pi_+-1}\widetilde{\ell}(\cdot)}{2}.$$

□

C. Discussion on Variance of Risk Estimator

C.1. Proof of Lemma 2

The statement can be simply confirmed as follows:

$$\mathop{\mathbb{E}}_{(X,X')\sim p_\mathrm{S}}\left[\frac{\mathcal{L}_{\mathrm{S},\ell}(f(X))+\mathcal{L}_{\mathrm{S},\ell}(f(X'))}{2}\right] = \mathop{\mathbb{E}}_{X\sim\widetilde{p}_\mathrm{S}}\left[\frac{\mathcal{L}_{\mathrm{S},\ell}(f(X))}{2}\right] + \mathop{\mathbb{E}}_{X'\sim\widetilde{p}_\mathrm{S}}\left[\frac{\mathcal{L}_{\mathrm{S},\ell}(f(X'))}{2}\right]$$

$$
\begin{aligned}
&= \mathop{\mathbb{E}}_{X\sim\widetilde{p}_{\mathrm{S}}}\left[\mathcal{L}_{\mathrm{S},\ell}(f(X))\right] \\
&= \mathop{\mathbb{E}}_{X\sim\widetilde{p}_{\mathrm{S}}}\left[\alpha\mathcal{L}_{\mathrm{S},\ell}(f(X)) + (1-\alpha)\mathcal{L}_{\mathrm{S},\ell}(f(X))\right] \\
&= \mathop{\mathbb{E}}_{X\sim\widetilde{p}_{\mathrm{S}}}\left[\alpha\mathcal{L}_{\mathrm{S},\ell}(f(X))\right] + \mathop{\mathbb{E}}_{X'\sim\widetilde{p}_{\mathrm{S}}}\left[(1-\alpha)\mathcal{L}_{\mathrm{S},\ell}(f(X'))\right] \\
&= \mathop{\mathbb{E}}_{(X,X')\sim p_{\mathrm{S}}}\left[\alpha\mathcal{L}_{\mathrm{S},\ell}(f(X)) + (1-\alpha)\mathcal{L}_{\mathrm{S},\ell}(f(X'))\right].
\end{aligned}
$$

□

C.2. Proof of Theorem 2

We show Eq. (8) is the variance minimizer of

$$
S(f;\alpha) \triangleq \frac{1}{n_{\mathrm{S}}}\sum_{i=1}^{n_{\mathrm{S}}}\left\{\alpha\mathcal{L}_{\mathrm{S},\ell}(f(\boldsymbol{x}_{\mathrm{S},i})) + (1-\alpha)\mathcal{L}_{\mathrm{S},\ell}(f(\boldsymbol{x}'_{\mathrm{S},i}))\right\},
$$

with respect to $\alpha \in \mathbb{R}$. Let $\mu_1 \triangleq \mathbb{E}_{\{(\boldsymbol{x}_{\mathrm{S},i},\boldsymbol{x}'_{\mathrm{S},i})\}\sim p_{\mathrm{S}}}[S(f;\alpha)]$ and

$$
\begin{aligned}
\widetilde{\mu}_1 &\triangleq \mathop{\mathbb{E}}_{\{\boldsymbol{x}_{\mathrm{S},i}\}\sim\widetilde{p}_{\mathrm{S}}}\left[\frac{1}{n_{\mathrm{S}}}\sum_{i=1}^{n_{\mathrm{S}}}\mathcal{L}_{\mathrm{S},\ell}(f(\boldsymbol{x}_{\mathrm{S},i}))\right] = \mathop{\mathbb{E}}_{\{\boldsymbol{x}'_{\mathrm{S},i}\}\sim\widetilde{p}_{\mathrm{S}}}\left[\frac{1}{n_{\mathrm{S}}}\sum_{i=1}^{n_{\mathrm{S}}}\mathcal{L}_{\mathrm{S},\ell}(f(\boldsymbol{x}'_{\mathrm{S},i}))\right], \\
\widetilde{\mu}_2 &\triangleq \mathop{\mathbb{E}}_{\{\boldsymbol{x}_{\mathrm{S},i}\}\sim\widetilde{p}_{\mathrm{S}}}\left[\left(\frac{1}{n_{\mathrm{S}}}\sum_{i=1}^{n_{\mathrm{S}}}\mathcal{L}_{\mathrm{S},\ell}(f(\boldsymbol{x}_{\mathrm{S},i}))\right)^2\right] = \mathop{\mathbb{E}}_{\{\boldsymbol{x}'_{\mathrm{S},i}\}\sim\widetilde{p}_{\mathrm{S}}}\left[\left(\frac{1}{n_{\mathrm{S}}}\sum_{i=1}^{n_{\mathrm{S}}}\mathcal{L}_{\mathrm{S},\ell}(f(\boldsymbol{x}'_{\mathrm{S},i}))\right)^2\right].
\end{aligned}
$$

Then,

$$
\begin{aligned}
\mathrm{Var}(S(f;\alpha)) &= \mathop{\mathbb{E}}_{\{(\boldsymbol{x}_{\mathrm{S},i},\boldsymbol{x}'_{\mathrm{S},i})\}\sim p_{\mathrm{S}}}\left[(S(f;\alpha)-\mu_1)^2\right] \\
&= \mathop{\mathbb{E}}_{\{(\boldsymbol{x}_{\mathrm{S},i},\boldsymbol{x}'_{\mathrm{S},i})\}\sim p_{\mathrm{S}}}\left[S(f;\alpha)^2\right] - \mu_1^2 \\
&= \alpha^2 \mathop{\mathbb{E}}_{\{\boldsymbol{x}_{\mathrm{S},i}\}\sim\widetilde{p}_{\mathrm{S}}}\left[\left(\frac{1}{n_{\mathrm{S}}}\sum_{i=1}^{n_{\mathrm{S}}}\mathcal{L}_{\mathrm{S},\ell}(f(\boldsymbol{x}_{\mathrm{S},i}))\right)^2\right] \\
&\quad + 2\alpha(1-\alpha)\mathop{\mathbb{E}}_{\{\boldsymbol{x}_{\mathrm{S},i}\}\sim\widetilde{p}_{\mathrm{S}}}\left[\frac{1}{n_{\mathrm{S}}}\sum_{i=1}^{n_{\mathrm{S}}}\mathcal{L}_{\mathrm{S},\ell}(f(\boldsymbol{x}_{\mathrm{S},i}))\right]\mathop{\mathbb{E}}_{\{\boldsymbol{x}'_{\mathrm{S},i}\}\sim\widetilde{p}_{\mathrm{S}}}\left[\frac{1}{n_{\mathrm{S}}}\sum_{i=1}^{n_{\mathrm{S}}}\mathcal{L}_{\mathrm{S},\ell}(f(\boldsymbol{x}'_{\mathrm{S},i}))\right] \\
&\quad + (1-\alpha)^2\mathop{\mathbb{E}}_{\{\boldsymbol{x}'_{\mathrm{S},i}\}\sim\widetilde{p}_{\mathrm{S}}}\left[\left(\frac{1}{n_{\mathrm{S}}}\sum_{i=1}^{n_{\mathrm{S}}}\mathcal{L}_{\mathrm{S},\ell}(f(\boldsymbol{x}'_{\mathrm{S},i}))\right)^2\right] - \mu_1^2 \\
&= \widetilde{\mu}_2\alpha^2 + 2\widetilde{\mu}_1^2\alpha(1-\alpha) + \widetilde{\mu}_2(1-\alpha)^2 - \mu_1^2 \\
&= 2(\widetilde{\mu}_2 - \widetilde{\mu}_1^2)\left\{\left(\alpha - \frac{1}{2}\right)^2 - \frac{1}{4}\right\} + \widetilde{\mu}_2 - \mu_1^2.
\end{aligned}
$$

Since $\widetilde{\mu}_2 - \widetilde{\mu}_1^2$ is the variance of $\frac{1}{n_{\mathrm{S}}}\sum_i \mathcal{L}_{\mathrm{S},\ell}(f(\boldsymbol{x}_{\mathrm{S},i}))$, $\widetilde{\mu}_2 - \widetilde{\mu}_1^2 \geq 0$. Thus, $\mathrm{Var}(S(f;\alpha))$ is minimized when $\alpha = \frac{1}{2}$. □

D. Proof of Theorem 3

Since ℓ is a twice differentiable margin loss, there is a twice differentiable function $\psi : \mathbb{R} \to \mathbb{R}_+$ such that $\ell(z,t) = \psi(tz)$. Taking the derivative of

$$
\widehat{J}_\ell(\boldsymbol{w}) = \frac{\lambda}{2}\boldsymbol{w}^\top\boldsymbol{w} - \frac{\pi_{\mathrm{S}}}{2n_{\mathrm{S}}(2\pi_+ - 1)}\sum_{i=1}^{2n_{\mathrm{S}}}\boldsymbol{w}^\top\boldsymbol{\phi}(\widetilde{\boldsymbol{x}}_{\mathrm{S},i})
$$

$$+ \frac{1}{n_{\mathrm{U}}(2\pi_+ - 1)} \sum_{i=1}^{n_{\mathrm{U}}} \left\{-\pi_- \ell(\boldsymbol{w}^\top \boldsymbol{\phi}(\boldsymbol{x}_{\mathrm{U},i}), +1) + \pi_+ \ell(\boldsymbol{w}^\top \boldsymbol{\phi}(\boldsymbol{x}_{\mathrm{U},i}), -1)\right\}$$

with respect to $\boldsymbol{w}$,

$$\frac{\partial}{\partial \boldsymbol{w}} \widehat{J}_\ell(\boldsymbol{w}) = \lambda \boldsymbol{w} - \frac{\pi_{\mathrm{S}}}{2n_{\mathrm{S}}(2\pi_+ - 1)} \sum_{i=1}^{2n_{\mathrm{S}}} \boldsymbol{\phi}(\widetilde{\boldsymbol{x}}_{\mathrm{S},i}) + \frac{1}{n_{\mathrm{U}}(2\pi_+ - 1)} \sum_{i=1}^{n_{\mathrm{U}}} \left\{-\pi_- \frac{\partial \ell(\xi_i, +1)}{\partial \xi_i} + \pi_+ \frac{\partial \ell(\xi_i, -1)}{\partial \xi_i}\right\} \boldsymbol{\phi}(\boldsymbol{x}_{\mathrm{U},i}),$$

where $\xi_i \triangleq \boldsymbol{w}^\top \boldsymbol{\phi}(\boldsymbol{x}_{\mathrm{U},i})$. Here, the second-order derivative of $\ell(z,t)$ with respect to z is

$$\frac{\partial^2 \ell(z,t)}{\partial z^2} = \frac{\partial^2 \psi(tz)}{\partial z^2} = \frac{\partial}{\partial z}\left(t \frac{\partial \psi(\xi)}{\partial \xi}\right) = t^2 \frac{\partial^2 \psi(\xi)}{\partial \xi^2} = \frac{\partial^2 \psi(\xi)}{\partial \xi^2},$$

where $\xi = tz$ is employed in the second equality and $t \in \{+1, -1\}$ is employed in the last equality. Thus the Hessian of $\widehat{J}_\ell$ is

$$\begin{aligned}
\boldsymbol{H} \widehat{J}_\ell(\boldsymbol{w}) &= \lambda I + \frac{1}{n_{\mathrm{U}}(2\pi_+ - 1)} \sum_{i=1}^{n_{\mathrm{U}}} \left\{-\pi_- \frac{\partial}{\partial \boldsymbol{w}} \frac{\partial \ell(\xi_i, +1)}{\partial \xi_i} + \pi_+ \frac{\partial}{\partial \boldsymbol{w}} \frac{\partial \ell(\xi_i, -1)}{\partial \xi_i}\right\} \boldsymbol{\phi}(\boldsymbol{x}_{\mathrm{U},i})^\top \\
&= \lambda I + \frac{1}{n_{\mathrm{U}}(2\pi_+ - 1)} \sum_{i=1}^{n_{\mathrm{U}}} \left\{-\pi_- \frac{\partial^2 \ell(\xi_i, +1)}{\partial \xi_i^2} \frac{\partial \xi_i}{\partial \boldsymbol{w}} + \pi_+ \frac{\partial^2 \ell(\xi_i, -1)}{\partial \xi_i^2} \frac{\partial \xi_i}{\partial \boldsymbol{w}}\right\} \boldsymbol{\phi}(\boldsymbol{x}_{\mathrm{U},i})^\top \\
&= \lambda I + \frac{1}{n_{\mathrm{U}}(2\pi_+ - 1)} \sum_{i=1}^{n_{\mathrm{U}}} \left\{-\pi_- \frac{\partial^2 \ell(\xi_i, +1)}{\partial \xi_i^2} + \pi_+ \frac{\partial^2 \ell(\xi_i, -1)}{\partial \xi_i^2}\right\} \boldsymbol{\phi}(\boldsymbol{x}_{\mathrm{U},i}) \boldsymbol{\phi}(\boldsymbol{x}_{\mathrm{U},i})^\top \\
&= \lambda I + \frac{1}{n_{\mathrm{U}}(2\pi_+ - 1)} \sum_{i=1}^{n_{\mathrm{U}}} (\pi_+ - \pi_-) \frac{\partial^2 \psi(\xi)}{\partial \xi^2} \boldsymbol{\phi}(\boldsymbol{x}_{\mathrm{U},i}) \boldsymbol{\phi}(\boldsymbol{x}_{\mathrm{U},i})^\top \\
&= \lambda I + \frac{1}{n_{\mathrm{U}}} \frac{\partial^2 \psi(\xi)}{\partial \xi^2} \sum_{i=1}^{n_{\mathrm{U}}} \boldsymbol{\phi}(\boldsymbol{x}_{\mathrm{U},i}) \boldsymbol{\phi}(\boldsymbol{x}_{\mathrm{U},i})^\top \\
&\succeq 0,
\end{aligned}$$

where $A \succeq 0$ means that a matrix A is positive semidefinite. Positive semidefiniteness of $\boldsymbol{H} \widehat{J}_\ell(\boldsymbol{w})$ follows from $\frac{\partial^2 \psi(\xi)}{\partial \xi^2} \geq 0$ ($\because \ell$ is convex) and $\boldsymbol{\phi}(\boldsymbol{x}_{\mathrm{U},i}) \boldsymbol{\phi}(\boldsymbol{x}_{\mathrm{U},i})^\top \succeq 0$. Thus $\widehat{J}_\ell(\boldsymbol{w})$ is convex. □

E. Derivation of Optimization Problems

E.1. Squared Loss

First, substituting the linear-in-parameter model $f(\boldsymbol{x}) = \boldsymbol{w}^\top \boldsymbol{\phi}(\boldsymbol{x})$ and the squared loss $\ell_{\mathrm{SQ}}(z,t) = \frac{1}{4}(tz-1)^2$ into Eq. (10), we obtain the following objective function:

$$\begin{aligned}
\widehat{J}_{\mathrm{SQ}}(\boldsymbol{w}) &= \frac{\pi_{\mathrm{S}}}{2(2\pi_+ - 1)n_{\mathrm{S}}} \sum_{i=1}^{2n_{\mathrm{S}}} \frac{(\boldsymbol{w}^\top \boldsymbol{\phi}(\widetilde{\boldsymbol{x}}_{\mathrm{S},i}) - 1)^2 - (\boldsymbol{w}^\top \boldsymbol{\phi}(\widetilde{\boldsymbol{x}}_{\mathrm{S},i}) + 1)^2}{4} \\
&\quad + \frac{1}{n_{\mathrm{U}}} \sum_{i=1}^{n_{\mathrm{U}}} \frac{-\pi_- \cdot \frac{1}{4}(\boldsymbol{w}^\top \boldsymbol{\phi}(\boldsymbol{x}_{\mathrm{U},i}) - 1)^2 + \pi_+ \cdot \frac{1}{4}(\boldsymbol{w}^\top \boldsymbol{\phi}(\boldsymbol{x}_{\mathrm{U},i}) + 1)^2}{2\pi_+ - 1} \\
&\quad + \frac{\lambda}{2}\|\boldsymbol{w}\|^2 \\
&= \frac{1}{2\pi_+ - 1} \left\{-\frac{\pi_{\mathrm{S}}}{2n_{\mathrm{S}}} \sum_{i=1}^{2n_{\mathrm{S}}} \boldsymbol{w}^\top \boldsymbol{\phi}(\widetilde{\boldsymbol{x}}_{\mathrm{S},i}) + \frac{1}{4n_{\mathrm{U}}} \sum_{i=1}^{n_{\mathrm{U}}} \left\{(2\pi_+ - 1)(\boldsymbol{w}^\top \boldsymbol{\phi}(\boldsymbol{x}_{\mathrm{U},i}) \boldsymbol{\phi}(\boldsymbol{x}_{\mathrm{U},i})^\top \boldsymbol{w} + 1) + 2\boldsymbol{\phi}(\boldsymbol{x}_{\mathrm{U},i})^\top \boldsymbol{w}\right\}\right\} \\
&\quad + \frac{\lambda}{2}\|\boldsymbol{w}\|^2
\end{aligned}$$

$$= \boldsymbol{w}^\top \left(\frac{1}{4n_{\mathrm{U}}} X_{\mathrm{U}}^\top X_{\mathrm{U}} + \frac{\lambda}{2} I \right) \boldsymbol{w} + \frac{1}{2\pi_+ - 1} \left(-\frac{\pi_{\mathrm{S}}}{2n_{\mathrm{S}}} \mathbf{1}^\top X_{\mathrm{S}} + \frac{1}{2n_{\mathrm{U}}} \mathbf{1}^\top X_{\mathrm{U}} \right) \boldsymbol{w}.$$

Taking the derivative with respect to $\boldsymbol{w}$,

$$\frac{\partial}{\partial \boldsymbol{w}} \widehat{J}_{\mathrm{SQ}}(\boldsymbol{w}) = \frac{1}{2n_{\mathrm{U}}} \left(X_{\mathrm{U}}^\top X_{\mathrm{U}} + 2n_{\mathrm{U}} \lambda I \right) \boldsymbol{w} - \frac{1}{2\pi_+ - 1} \left(\frac{\pi_{\mathrm{S}}}{2n_{\mathrm{S}}} X_{\mathrm{S}}^\top \mathbf{1} - \frac{1}{2n_{\mathrm{U}}} X_{\mathrm{U}}^\top \mathbf{1} \right).$$

Solving $\frac{\partial}{\partial \boldsymbol{w}} \widehat{J}_{\mathrm{SQ}}(\boldsymbol{w}) = 0$, we obtain the analytical solution:

$$\boldsymbol{w} = \frac{n_{\mathrm{U}}}{2\pi_+ - 1} (X_{\mathrm{U}}^\top X_{\mathrm{U}} + 2n_{\mathrm{U}} \lambda I)^{-1} \left(\frac{\pi_{\mathrm{S}}}{n_{\mathrm{S}}} X_{\mathrm{S}}^\top \mathbf{1} - \frac{1}{n_{\mathrm{U}}} X_{\mathrm{U}}^\top \mathbf{1} \right).$$

E.2. Double-Hinge Loss

Using the double-hinge loss $\ell_{\mathrm{DH}}(z, t) = \max(-tz, \max(0, \frac{1}{2} - \frac{1}{2} tz))$, we obtain the following objective function:

$$\begin{aligned} \widehat{J}_{\mathrm{DH}}(\boldsymbol{w}) = & -\frac{\pi_{\mathrm{S}}}{2n_{\mathrm{S}}(2\pi_+ - 1)} \sum_{i=1}^{2n_{\mathrm{S}}} \boldsymbol{w}^\top \boldsymbol{\phi}(\widetilde{\boldsymbol{x}}_{\mathrm{S},i}) \\ & + \frac{1}{n_{\mathrm{U}}(2\pi_+ - 1)} \sum_{i=1}^{n_{\mathrm{U}}} \left\{ -\pi_- \ell_{\mathrm{DH}}(\boldsymbol{w}^\top \boldsymbol{\phi}(\boldsymbol{x}_{\mathrm{U},i})) + \pi_+ \ell_{\mathrm{DH}}(-\boldsymbol{w}^\top \boldsymbol{\phi}(\boldsymbol{x}_{\mathrm{U},i})) \right\} \\ & + \frac{\lambda}{2} \boldsymbol{w}^\top \boldsymbol{w}. \end{aligned}$$

Using slack variables $\boldsymbol{\xi}, \boldsymbol{\eta} \in \mathbb{R}^{n_{\mathrm{U}}}$, the objective function can be rewritten into the following optimization problem:

$$\min_{\boldsymbol{w}, \boldsymbol{\xi}, \boldsymbol{\eta}} -\frac{\pi_{\mathrm{S}}}{2n_{\mathrm{S}}(2\pi_+ - 1)} \mathbf{1}^\top X_{\mathrm{S}} \boldsymbol{w} - \frac{\pi_-}{n_{\mathrm{S}}(2\pi_+ - 1)} \mathbf{1}^\top \boldsymbol{\xi} + \frac{\pi_+}{n_{\mathrm{U}}(2\pi_+ - 1)} \mathbf{1}^\top \boldsymbol{\eta} + \frac{\lambda}{2} \boldsymbol{w}^\top \boldsymbol{w}$$

$$\begin{aligned} \text{s.t.} \quad & \boldsymbol{\xi} \geq \mathbf{0}, \quad \boldsymbol{\xi} \geq \frac{1}{2} \mathbf{1} + \frac{1}{2} X_{\mathrm{U}} \boldsymbol{w}, \quad \boldsymbol{\xi} \geq X_{\mathrm{U}} \boldsymbol{w}, \\ & \boldsymbol{\eta} \geq \mathbf{0}, \quad \boldsymbol{\eta} \geq \frac{1}{2} \mathbf{1} - \frac{1}{2} X_{\mathrm{U}} \boldsymbol{w}, \quad \boldsymbol{\eta} \geq -X_{\mathrm{U}} \boldsymbol{w}, \end{aligned}$$

where $\geq$ for vectors denotes the element-wise inequality.

Below, we rewrite the optimization problem into the standard QP form. Let $\boldsymbol{\gamma} \triangleq [\boldsymbol{w}^\top \ \boldsymbol{\xi}^\top \ \boldsymbol{\eta}^\top]^\top \in \mathbb{R}^{d+2n_{\mathrm{U}}}$ be a objective variable and

$$P \triangleq \begin{bmatrix} \lambda I_d & O_{d,n_{\mathrm{U}}} & O_{d,n_{\mathrm{U}}} \\ O_{n_{\mathrm{U}},d} & O_{n_{\mathrm{U}},n_{\mathrm{U}}} & O_{n_{\mathrm{U}},n_{\mathrm{U}}} \\ O_{n_{\mathrm{U}},d} & O_{n_{\mathrm{U}},n_{\mathrm{U}}} & O_{n_{\mathrm{U}},n_{\mathrm{U}}} \end{bmatrix}, \qquad \boldsymbol{q} \triangleq \begin{bmatrix} -\frac{\pi_{\mathrm{S}}}{2n_{\mathrm{S}}(2\pi_+ - 1)} X_{\mathrm{S}}^\top \mathbf{1}_d \\ -\frac{\pi_-}{n_{\mathrm{U}}(2\pi_+ - 1)} \mathbf{1}_{n_{\mathrm{U}}} \\ \frac{\pi_+}{n_{\mathrm{U}}(2\pi_+ - 1)} \mathbf{1}_{n_{\mathrm{U}}} \end{bmatrix}$$

$$G \triangleq \begin{bmatrix} O_{n_{\mathrm{U}},d} & -I_{n_{\mathrm{U}}} & O_{n_{\mathrm{U}},n_{\mathrm{U}}} \\ \frac{1}{2} X_{\mathrm{U}} & -I_{n_{\mathrm{U}}} & O_{n_{\mathrm{U}},n_{\mathrm{U}}} \\ X_{\mathrm{U}} & -I_{n_{\mathrm{U}}} & O_{n_{\mathrm{U}},n_{\mathrm{U}}} \\ O_{n_{\mathrm{U}},d} & O_{n_{\mathrm{U}},n_{\mathrm{U}}} & -I_{n_{\mathrm{U}}} \\ -\frac{1}{2} X_{\mathrm{U}} & O_{n_{\mathrm{U}},n_{\mathrm{U}}} & -I_{n_{\mathrm{U}}} \\ -X_{\mathrm{U}} & O_{n_{\mathrm{U}},n_{\mathrm{U}}} & -I_{n_{\mathrm{U}}} \end{bmatrix}, \qquad \boldsymbol{h} \triangleq \begin{bmatrix} \mathbf{0}_{n_{\mathrm{U}}} \\ -\frac{1}{2} \mathbf{1}_{n_{\mathrm{U}}} \\ \mathbf{0}_{n_{\mathrm{U}}} \\ \mathbf{0}_{n_{\mathrm{U}}} \\ -\frac{1}{2} \mathbf{1}_{n_{\mathrm{U}}} \\ \mathbf{0}_{n_{\mathrm{U}}} \end{bmatrix},$$

where I_k means $k \times k$ identity matrix, $O_{k,l}$ means $k \times l$ all-zero matrix, $\mathbf{1}_k$ is k-dimensional all-one vector, and $\mathbf{0}_k$ is k-dimensional all-zero vector. Then the optimization problem is

$$\min_{\boldsymbol{\gamma}} \frac{1}{2} \boldsymbol{\gamma}^\top P \boldsymbol{\gamma} + \boldsymbol{q}^\top \boldsymbol{\gamma} \qquad \text{s.t.} \quad G\boldsymbol{\gamma} \leq \boldsymbol{h},$$

which is the standard form of QP.

F. Proof of Theorem 4

First, we derive the next risk expression for convenience.

Lemma 4. *Given any function $f : \mathcal{X} \to \mathbb{R}$, let $R_{\widetilde{\mathrm{S}}\mathrm{U},\ell}(f)$ be*

$$R_{\widetilde{\mathrm{S}}\mathrm{U},\ell}(f) = \pi_{\mathrm{S}} \mathop{\mathbb{E}}_{X \sim \widetilde{p}_{\mathrm{S}}} [\mathcal{L}_{\mathrm{S},\ell}(f(X))] + \mathop{\mathbb{E}}_{X \sim p} [\mathcal{L}_{\mathrm{U},\ell}(f(X))],$$

then $R_{\mathrm{SU},\ell}(f) = R_{\widetilde{\mathrm{S}}\mathrm{U},\ell}(f)$.

Proof. The first term on the RHS of Eq. (5) can be transformed as follows:

$$\begin{aligned}\pi_{\mathrm{S}} \mathop{\mathbb{E}}_{(X,X') \sim p_{\mathrm{S}}} \left[\frac{\mathcal{L}_{\mathrm{S},\ell}(f(X)) + \mathcal{L}_{\mathrm{S},\ell}(f(X'))}{2} \right] &= \pi_{\mathrm{S}} \left\{ \frac{1}{2} \mathop{\mathbb{E}}_{X \sim \widetilde{p}_{\mathrm{S}}} [\mathcal{L}_{\mathrm{S},\ell}(f(X))] + \frac{1}{2} \mathop{\mathbb{E}}_{X' \sim \widetilde{p}_{\mathrm{S}}} [\mathcal{L}_{\mathrm{S},\ell}(f(X'))] \right\} \\ &= \pi_{\mathrm{S}} \mathop{\mathbb{E}}_{X \sim \widetilde{p}_{\mathrm{S}}} [\mathcal{L}_{\mathrm{S},\ell}(f(X))].\end{aligned}$$

□

Next, we show the uniform deviation bound, which is useful to derive estimation error bounds. The proof can be found in the textbooks such as Mohri et al. (2012) (Theorem 3.1).

Lemma 5. *Let Z be a random variable drawn from a probability distribution with density μ, $\mathcal{H} = \{h : \mathcal{Z} \to [0, M]\}$ ($M > 0$) be a class of measurable functions, $\{z_i\}_{i=1}^n$ be i.i.d. samples drawn from the distribution with density μ. Then, with probability at least $1 - \delta$,*

$$\sup_{h \in \mathcal{H}} \left| \mathop{\mathbb{E}}_{Z \sim \mu} [h(Z)] - \frac{1}{n} \sum_{i=1}^{n} h(z_i) \right| \leq 2\mathfrak{R}(\mathcal{H}; n, \mu) + \sqrt{\frac{M^2 \log \frac{2}{\delta}}{2n}}.$$

Let us begin with the estimation error $R(\widehat{f}) - R(f^*)$. For convenience, let

$$\begin{aligned} R_{\widetilde{\mathrm{S}},\ell}(f) &\triangleq \mathop{\mathbb{E}}_{X \sim \widetilde{p}_{\mathrm{S}}} [\mathcal{L}_{\mathrm{S},\ell}(f(X))], \qquad & \widehat{R}_{\widetilde{\mathrm{S}},\ell}(f) &\triangleq \frac{1}{2n_{\mathrm{S}}} \sum_{i=1}^{2n_{\mathrm{S}}} \mathcal{L}_{\mathrm{S},\ell}(f(\widetilde{\boldsymbol{x}}_{\mathrm{S},i})), \\ R_{\mathrm{U},\ell}(f) &\triangleq \mathop{\mathbb{E}}_{X \sim p} [\mathcal{L}_{\mathrm{U},\ell}(f(X))], \qquad & \widehat{R}_{\mathrm{U},\ell}(f) &\triangleq \frac{1}{n_{\mathrm{U}}} \sum_{i=1}^{n_{\mathrm{U}}} \mathcal{L}_{\mathrm{U},\ell}(f(\boldsymbol{x}_{\mathrm{U},i})), \\ \widehat{R}_{\widetilde{\mathrm{S}}\mathrm{U},\ell}(f) &\triangleq \pi_{\mathrm{S}} \widehat{R}_{\widetilde{\mathrm{S}},\ell}(f) + \widehat{R}_{\mathrm{U},\ell}(f). \end{aligned}$$

Note that

$$\widehat{R}_{\mathrm{SU},\ell}(f) = \widehat{R}_{\widetilde{\mathrm{S}}\mathrm{U},\ell}(f) \tag{F.1}$$

by Eq. (5). Then,

$$\begin{aligned} R(\widehat{f}) - R(f^*) &= R_{\mathrm{SU},\ell}(\widehat{f}) - R_{\mathrm{SU},\ell}(f^*) && (\because \text{Theorem 1}) \\ &= (R_{\mathrm{SU},\ell}(\widehat{f}) - \widehat{R}_{\mathrm{SU},\ell}(\widehat{f})) + (\widehat{R}_{\mathrm{SU},\ell}(\widehat{f}) - \widehat{R}_{\mathrm{SU},\ell}(f^*)) \\ &\quad + (\widehat{R}_{\mathrm{SU},\ell}(f^*) - R_{\mathrm{SU},\ell}(f^*)) \\ &\leq (R_{\mathrm{SU},\ell}(\widehat{f}) - \widehat{R}_{\mathrm{SU},\ell}(\widehat{f})) + 0 + (\widehat{R}_{\mathrm{SU},\ell}(f^*) - R_{\mathrm{SU},\ell}(f^*)) && (\because \text{by the definition of } f^* \text{ and } \widehat{f}) \\ &\leq 2 \sup_{f \in \mathcal{F}} \left| R_{\mathrm{SU},\ell}(f) - \widehat{R}_{\mathrm{SU},\ell}(f) \right| \\ &= 2 \sup_{f \in \mathcal{F}} \left| R_{\widetilde{\mathrm{S}}\mathrm{U},\ell}(f) - \widehat{R}_{\widetilde{\mathrm{S}}\mathrm{U},\ell}(f) \right| && (\because \text{Lemma 4 and Eq. (F.1)}) \\ &\leq 2\pi_{\mathrm{S}} \sup_{f \in \mathcal{F}} \left| R_{\widetilde{\mathrm{S}},\ell}(f) - \widehat{R}_{\widetilde{\mathrm{S}},\ell}(f) \right| + 2 \sup_{f \in \mathcal{F}} \left| R_{\mathrm{U},\ell}(f) - \widehat{R}_{\mathrm{U},\ell}(f) \right| && (\because \text{subadditivity of } \sup). \end{aligned} \tag{F.2}$$

Each term in the last line is bounded in next two lemmas with probability at least $1 - \frac{\delta}{2}$.

Lemma 6. *Assume the loss function ℓ is ρ-Lipschitz with respect to the first argument ($0 < \rho < \infty$), and all functions in the model class $\mathcal{F}$ are bounded, i.e., there exists a constant C_b such that $\|f\|_\infty \le C_b$ for any $f \in \mathcal{F}$. Let $C_\ell \triangleq \sup_{t\in\{\pm1\}} \ell(C_b, t)$. For any $\delta > 0$, with probability at least $1 - \frac{\delta}{2}$,*

$$\sup_{f\in\mathcal{F}} \left| R_{\widetilde{\mathrm{S}},\ell}(f) - \widehat{R}_{\widetilde{\mathrm{S}},\ell}(f) \right| \le \frac{4\rho C_{\mathcal{F}} + \sqrt{2C_\ell^2 \log\frac{4}{\delta}}}{|2\pi_+ - 1|\sqrt{2n_\mathrm{S}}}$$

Proof. By Lemma 5,

$$\begin{aligned}
\sup_{f\in\mathcal{F}} \left| R_{\widetilde{\mathrm{S}},\ell}(f) - \widehat{R}_{\widetilde{\mathrm{S}},\ell}(f) \right| &= \frac{1}{|2\pi_+ - 1|} \sup_{f\in\mathcal{F}} \left| \mathop{\mathbb{E}}_{X\sim\widetilde{p}_\mathrm{S}} \left[\widetilde{\ell}(f(X)) \right] - \frac{1}{2n_\mathrm{S}} \sum_{i=1}^{2n_\mathrm{S}} \widetilde{\ell}(f(\widetilde{\boldsymbol{x}}_{\mathrm{S},i})) \right| \\
&\le \frac{1}{|2\pi_+ - 1|} \left\{ \sup_{f\in\mathcal{F}} \left| \mathop{\mathbb{E}}_{X\sim\widetilde{p}_\mathrm{S}} [\ell(f(X), +1)] - \frac{1}{2n_\mathrm{S}} \sum_{i=1}^{2n_\mathrm{S}} \ell(f(\widetilde{\boldsymbol{x}}_{\mathrm{S},i}), +1) \right| \right. \\
&\quad \left. + \sup_{f\in\mathcal{F}} \left| \mathop{\mathbb{E}}_{X\sim\widetilde{p}_\mathrm{S}} [\ell(f(X), -1)] - \frac{1}{2n_\mathrm{S}} \sum_{i=1}^{2n_\mathrm{S}} \ell(f(\widetilde{\boldsymbol{x}}_{\mathrm{S},i}), -1) \right| \right\} \\
&\le \frac{1}{|2\pi_+ - 1|} \left\{ 4\mathfrak{R}(\ell\circ\mathcal{F}; 2n_\mathrm{S}, p_\mathrm{S}) + \sqrt{\frac{2C_\ell^2 \log\frac{4}{\delta}}{2n_\mathrm{S}}} \right\},
\end{aligned}$$

where $\ell \circ \mathcal{F}$ in the last line means $\{\ell \circ f \mid f \in \mathcal{F}\}$. The last inequality holds from Lemma 5. By Talagrand's lemma (Lemma 4.2 in Mohri et al. (2012)),

$$\mathfrak{R}(\ell\circ\mathcal{F}; 2n_\mathrm{S}, p_\mathrm{S}) \le \rho\mathfrak{R}(\mathcal{F}; 2n_\mathrm{S}, p_\mathrm{S}).$$

Together with Eq. (13), we obtain

$$\begin{aligned}
\sup_{f\in\mathcal{F}} \left| R_{\widetilde{\mathrm{S}},\ell}(f) - \widehat{R}_{\widetilde{\mathrm{S}},\ell}(f) \right| &\le \frac{1}{|2\pi_+ - 1|} \left\{ 4\rho\frac{C_{\mathcal{F}}}{\sqrt{2n_\mathrm{S}}} + \sqrt{\frac{2C_\ell^2 \log\frac{4}{\delta}}{2n_\mathrm{S}}} \right\} \\
&= \frac{4\rho C_{\mathcal{F}} + \sqrt{2C_\ell^2 \log\frac{4}{\delta}}}{|2\pi_+ - 1|\sqrt{2n_\mathrm{S}}}.
\end{aligned}$$

□

Lemma 7. *Assume the loss function ℓ is ρ-Lipschitz with respect to the first argument ($0 < \rho < \infty$), and all functions in the model class $\mathcal{F}$ are bounded, i.e., there exists a constant C_b such that $\|f\|_\infty \le C_b$ for any $f \in \mathcal{F}$. Let $C_\ell \triangleq \sup_{t\in\{\pm1\}} \ell(C_b, t)$. For any $\delta > 0$, with probability at least $1 - \frac{\delta}{2}$,*

$$\sup_{f\in\mathcal{F}} \left| R_{\mathrm{U},\ell}(f) - \widehat{R}_{\mathrm{U},\ell}(f) \right| \le \frac{2\rho C_{\mathcal{F}} + \sqrt{\frac{1}{2}C_\ell^2 \log\frac{4}{\delta}}}{|2\pi_+ - 1|\sqrt{n_\mathrm{U}}}$$

Proof. This lemma can be proven similarly to Lemma 6. □

Combining Lemma 6, Lemma 7 and Eq. (F.2), Theorem 4 is proven. □