Supplementary material

Appendix A provides proofs of theorems stated in Sec 2, where we presented the proposed reject option models for the OOD setup and their optimal strategies. Appendix A is organized as follows:

- Appendix A.1. Proof of Theorem 1 providing an optimal strategy of the cost-based OOD model.
- Appendix A.2. Proof of Theorem 2 and Theorem 4 that claim that the Bayes ID classifier (4) is an optimal solution of the bounded TPR-FPR and the bounded Precision-Recall model, respectively. The proof of both theorems is the same, hence we put it to the same section.
- Appendix A.3. Proof of Theorem 3 providing a form of an optimal selective function under the bounded TPR-FPR model for an arbitrary fixed ID classifier.
- Appendix A.4. In this section, we characterize the form of $\tau(x)$ function, which defines the acceptance probability of boundary inputs $\mathcal{X}_{s(x)=\lambda} = \{x \in \mathcal{X} \mid s(x) = \lambda\}$ for the optimal selective function (9).
- Appendix A.5. In the case of finite input space, i.e. $|\mathcal{X}| < \infty$, we can find an optimal selective function under the bounded TRP-FPR model via Linear Programming described in this section.
- Appendix A.6. Proof of Theorem 5 providing a form of an optimal selective function under the Bounded Precision-Recall model for an arbitrary fixed ID classifier.

Appendix B provides supplementary material for Sec. 3. The evaluation curves obtained for the examplar methods on synthetic data are shown in Sec B.1. The algorithm to solve the problems (13) and (14) is discussed in Sec. B.2.

A Proofs of theorems from Sec 2

A.1 Proof of Theorem 1

Due to the additivity of the expected risk $R(h,c) = \mathbb{E}_{x,y\sim p(x,\bar{y})} \bar{\ell}(\bar{y},(h,c)(x))$, the optimal strategy minimizing the risk can be found for each input $x \in \mathcal{X}$ separately by solving

$$q^*(x) = \operatorname*{argmin}_{q\in\bar{\mathcal{Y}}} R_x(q)$$

where $R_x(q)$ is the partial risk defined as

$$R_x(q) = \sum_{\bar{y}\in\bar{\mathcal{Y}}} p(x,\bar{y})\,\bar{\ell}(\bar{y},q) = p_O(x)\pi\Big(\llbracket q=\emptyset\rrbracket\,\epsilon_3 + \llbracket q(x)\neq\emptyset\rrbracket\epsilon_2\Big) + (1-\pi)\sum_{y\in\mathcal{Y}} p_I(x,y)\Big(\llbracket q=\emptyset\rrbracket\epsilon_1 + \llbracket q\neq\emptyset\rrbracket\,\ell(y,q)\Big)$$

We can se that

$$\begin{aligned}
R_x(q=\text{reject}) &= p_O(x)\,\pi\,\epsilon_3 + P_I(x)\,(1-\pi)\,\epsilon_1\\
R_x(q\neq\text{reject}) &= p_O(x)\,\pi\,\epsilon_2 + (1-\pi)\sum_{y\in\mathcal{Y}} p_I(x,y)\,\ell(y,q)\\
\min_{q\in\mathcal{Y}} R_x(q) &= p_O(x)\,\pi\,\epsilon_2 + (1-\pi)\,p_I(x)\,r_B(x)\;,
\end{aligned}$$

where $r(x)$ is the minimal conditional risk

$$r_B(x) = \min_{\hat{y}\in\mathcal{Y}} \sum_{y\in\mathcal{Y}} p_I(y\mid x)\ell(y,\hat{y})$$

It is optimal to reject when

$$\begin{aligned}
0 &\leq \min_{q\in\mathcal{Y}} R_x(q) - R_x(q=\text{reject})\\
&= p_O(x)\,\pi\,\epsilon_2 + (1-\pi)\,p_I(x)\,r(x) - p_O(x)\,\pi\,\epsilon_3 - p_I(x)\,(1-\pi)\,\epsilon_1\\
&= p_O(x)\,\pi\,(\epsilon_2-\epsilon_3) + (1-\pi)\,p_I(x)\,(r_B(x)-\epsilon_1)\\
&= s(x)\;.
\end{aligned}$$

The inequality $0 \leq s(x)$ is equalivalent to

$$r_B(x) + (\varepsilon_2 - \varepsilon_3)\frac{\pi}{1-\pi} g(x) \geq \epsilon_1 \,.$$

In case that $\pi < 10$ and $\frac{K}{0} = \infty$, the optimal strategy then reads

$$q^* = \left\{ \begin{array}{rll} \text{reject} & \text{if} & s(x) \geq 0 \\ \underset{\hat{y} \in \mathcal{Y}}{\operatorname{argmin}} \sum_{y \in \mathcal{Y}} p_I(x,y) \ell(y,\hat{y}) & \text{if} & s(x) \leq 0 \end{array} \right.$$

Note that in the boundary case $s(x) = 0$ we can reject or accept arbitrarily without affecting the solution.

A.2 Proof of Theorem 2 and Theorem 4

The definition of h_B allows to derive $\mathrm{R}^{\mathrm{S}}(h_B, c) \leq \mathrm{R}^{\mathrm{S}}(h, c)$ as follows:

$$\begin{aligned}
\mathrm{R}^{\mathrm{S}}(h_B, c) &= \frac{1}{\phi(c)} \int\limits_{\mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y)\, \ell(y, h_B(x))\, c(x)\, dx \\
&= \frac{1}{\phi(c)} \int\limits_{\mathcal{X}} p(x) c(x) \left(\sum_{y \in \mathcal{Y}} p(y \,|\, x)\, \ell(y, h_B(x)) \right) dx \\
&\leq \frac{1}{\phi(c)} \int\limits_{\mathcal{X}} p(x) c(x) \left(\sum_{y \in \mathcal{Y}} p(y \,|\, x)\, \ell(y, h(x)) \right) dx \\
&= \frac{1}{\phi(c)} \int\limits_{\mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y)\, \ell(y, h(x))\, c(x)\, dx \\
&= \mathrm{R}^{\mathrm{S}}(h, c)\,.
\end{aligned}$$

A.3 Proof of Theorem 3

It is a direct consequence of the following theorem.

Theorem 6. *For any (h, c) optimal to (8), there exist real numbers λ, μ such that*

$$\begin{aligned}
\int\limits_{\mathcal{X}^<} p_I(x) c(x) dx &= \int\limits_{\mathcal{X}^<} p_I(x) dx \,, \\
\int\limits_{\mathcal{X}^>} p_I(x) c(x) dx &= 0 \,,
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{X}^< &= \{x \in \mathcal{X} \mid r(x) + \mu \frac{p_O(x)}{p_I(x)} < \lambda\} \,, \\
\mathcal{X}^> &= \{x \in \mathcal{X} \mid r(x) + \mu \frac{p_O(x)}{p_I(x)} > \lambda\} \,.
\end{aligned}$$

Proof. We first give a proof for countable sets $\mathcal{X}$, when integrals can be expressed as sums, then we present its generalization to arbitrary $\mathcal{X}$.

Assume $\mathcal{X}$ is countable and (h, c) is optimal to (8). Observe that we do not need to pay attention to those $x \in \mathcal{X}$ for which $p_I(x) = 0$ as they do not have any impact on the theorem statement. Denote

$$\begin{aligned}
\mathcal{X}^+ &= \{x \in \mathcal{X} \mid p_I(x) > 0\} \,, \\
\mathcal{X}_0 &= \left\{x \in \mathcal{X}^+ \mid c(x) = 0\right\} \,, \\
\mathcal{X}_1 &= \left\{x \in \mathcal{X}^+ \mid c(x) = 1\right\} \,, \\
\mathcal{X}_2 &= \left\{x \in \mathcal{X}^+ \mid 0 < c(x) < 1\right\} \,.
\end{aligned}$$

Let $P : \mathcal{X}^+ \to \mathbb{R}^2_+$ be a mapping such that $P(x) = \left(\frac{p_O(x)}{p_I(x)}, \frac{R(x)}{p_I(x)}\right)$, where

$$R(x) = \sum_{y \in \mathcal{Y}} p(x,y)\ell(y, h(x)) \,.$$

To confirm the existence of suitable λ, μ, it suffices to show that the sets

$$A_0 = \{P(x) \mid x \in \mathcal{X}_0\} \,, \tag{15}$$
$$A_1 = \{P(x) \mid x \in \mathcal{X}_1\} \,, \tag{16}$$
$$A_2 = \{P(x) \mid x \in \mathcal{X}_2\} \tag{17}$$

are "almost" linearly separable, i.e., there is a line L that includes A_2 and linearly separates the sets $A_0 \setminus L$, $A_1 \setminus L$. The existence of such L is ensured if

$$\dim \operatorname{span}\left((\operatorname{conv}(A_0) \cap \operatorname{conv}(A_1)) \cup A_2\right) < 2 \,, \tag{18}$$

where $\operatorname{conv}(\cdot)$ denotes the convex hull and $\operatorname{span}(\cdot)$ denotes the span of a set of vectors.

We will check validity of condition (18) by using the following two claims.

Claim 6.1. *Let $x_1, x_2 \in \mathcal{X}^+$, $r(x_1) > r(x_2)$, and $\frac{p_O(x_1)}{p_I(x_1)} \geq \frac{p_O(x_2)}{p_I(x_2)}$. Then, $x_1 \in \mathcal{X}_0$ or $x_2 \in \mathcal{X}_1$.*

Proof of the claim. By contradiction. Assume $c(x_1) > 0$ and $c(x_2) < 1$. Define a selective function c' which is identical to c up to $c'(x_1) = c(x_1) - \Delta$, $c'(x_2) = c(x_2) + \frac{p_I(x_1)}{p_I(x_2)}\Delta$, where

$$\Delta = \min\left\{c(x_1), \frac{p_I(x_2)}{p_I(x_1)}(1 - c(x_2))\right\} > 0 \,.$$

Now, observe that

$$\phi(c') - \phi(c) = -\Delta \cdot p_I(x_1) + \frac{p_I(x_1)}{p_I(x_2)}\Delta \cdot p_I(x_2) = 0 \,,$$

$$\rho(c') - \rho(c) = -\Delta \cdot p_O(x_1) + \frac{p_I(x_1)}{p_I(x_2)}\Delta \cdot p_O(x_2) \leq 0 \,, \quad \text{and}$$

$$\phi(c)\left(\mathrm{R}^{\mathrm{S}}(h,c') - \mathrm{R}^{\mathrm{S}}(h,c)\right) = -\Delta \cdot R(x_1) + \frac{p_I(x_1)}{p_I(x_2)}\Delta \cdot R(x_2) = \Delta \cdot p_I(x_1)(r(x_2) - r(x_1)) < 0$$

contradicts the optimality of (h, c). ■

Claim 6.2. *Let x_1, x_2, x_3 be elements of $\mathcal{X}^+$ such that the points $P_1 = P(x_1), P_2 = P(x_2), P_3 = P(x_3)$ are non-collinear and $\beta \cdot P_3 = \alpha_1 \cdot P_1 + \alpha_2 \cdot P_2$ holds for some $\alpha_1, \alpha_2, \beta \in \mathbb{R}_+$, where $\alpha_1 + \alpha_2 = 1$.*

- *If $\beta < 1$, then $x_3 \in \mathcal{X}_0$ or $\{x_1, x_2\} \cap \mathcal{X}_1 \neq \emptyset$.*
- *If $\beta > 1$, then $x_3 \in \mathcal{X}_1$ or $\{x_1, x_2\} \cap \mathcal{X}_0 \neq \emptyset$.*

Proof of the claim. We will give a proof for $\beta < 1$ and note that the steps for $\beta > 1$ are analogous.

By contradiction. Assume $c(x_1) < 1$, $c(x_2) < 1$, and $c(x_3) > 0$. To simplify the notation, for $i \in \{1, 2, 3\}$, let $p_i = p_I(x_i)$, $q_i = p_O(x_i)$, and $R_i = R(x_i)$.

Define a selective function c' which is identical to c up to

$$c'(x_1) = c(x_1) + \Delta \cdot \alpha_1 \frac{p_3}{p_1} \,,$$
$$c'(x_2) = c(x_2) + \Delta \cdot \alpha_2 \frac{p_3}{p_2} \,,$$
$$c'(x_3) = c(x_3) - \Delta \,,$$

where

$$\Delta = \min\left\{c(x_3), \frac{p_1}{\alpha_1 p_3}(1-c(x_1)), \frac{p_2}{\alpha_2 p_3}(1-c(x_2))\right\} > 0\,.$$

Observe that

$$\phi(c') - \phi(c) = \Delta\alpha_1 p_3 + \Delta\alpha_2 p_3 - \Delta p_3 = 0\,,$$

$$\rho(c') - \rho(c) = \Delta\alpha_1 \frac{p_3}{p_1} q_1 + \Delta\alpha_2 \frac{p_3}{p_2} q_2 - \Delta q_3 = \Delta(\beta-1) q_3 \leq 0\,, \quad \text{and}$$

$$\begin{aligned}\phi(c)\left(\mathrm{R}^{\mathrm{S}}(h,c') - \mathrm{R}^{\mathrm{S}}(h,c)\right) &= \Delta\alpha_1 \frac{p_3}{p_1} R_1 + \Delta\alpha_2 \frac{p_3}{p_2} R_2 - \Delta R_3 \\ &= \Delta p_3 \left(\alpha_1 \frac{R_1}{p_1} + \alpha_2 \frac{R_2}{p_2}\right) - \Delta R_3 = \Delta p_3 \beta \frac{R_3}{p_3} - \Delta R_3 \\ &= \Delta(\beta-1) R_3 < 0\end{aligned}$$

contradicts the optimality of c. ■

We are ready to confirm condition (18), this is done by analyzing the potential infeasible cases.

1. $\dim \operatorname{span}(\operatorname{conv}(A_0) \cap \operatorname{conv}(A_1)) = 2$. Then, there are $x_1, x_2, x_3, x_4 \in \mathcal{X}^+$ such that $P(x_1), P(x_2), P(x_3)$ are non-collinear, $P(x_4)$ is inside the triangle $P(x_1), P(x_2), P(x_3)$, and, either $x_1, x_2, x_3 \in \mathcal{X}_0$, $x_4 \in \mathcal{X}_1$, or $x_1, x_2, x_3 \in \mathcal{X}_1$, $x_4 \in \mathcal{X}_0$.

2. $\dim \operatorname{span}(A_2) = 2$. There are $x_1, x_2, x_3 \in \mathcal{X}_2$ such that $P(x_1), P(x_2), P(x_3)$ are non-collinear.

3. $\dim \operatorname{span}(\operatorname{conv}(A_0) \cap \operatorname{conv}(A_1)) = 1$ and $\dim \operatorname{span}((\operatorname{conv}(A_0) \cap \operatorname{conv}(A_1)) \cup A_2) = 2$. There are $x_1, x_2 \in \mathcal{X}_0$, $x_3, x_4 \in \mathcal{X}_1$, $x_5 \in \mathcal{X}_2$ such that points $P(x_1), P(x_3)$ lie on a half-line H_1, points $P(x_2), P(x_4)$ lie on a half-line H_2, where $H_1 \cap H_2 = \emptyset$ and $\operatorname{conv}(H_1 \cup H_2)$ is a line not containing $P(x_5)$.

4. $\dim \operatorname{span}(A_2) = 1$, and $\dim \operatorname{span}((\operatorname{conv}(A_0) \cap \operatorname{conv}(A_1)) \cup A_2) = 2$. There are $x_1 \in \mathcal{X}_0$, $x_2 \in \mathcal{X}_1$, $x_3, x_4 \in \mathcal{X}_2$ such that $P(x_3) \neq P(x_4)$, points $P(x_3), P(x_4)$ lie on a line L, and points $P(x_1)$, $P(x_2)$ lie in one half-plane of L, but not on L.

It is not difficult to check that all the listed points configurations always enable to select a subset of two or three points whose existence is ruled out by Claim 6.1 or Claim 6.2, respectively.

Consider now that $\mathcal{X}$ is an arbitrary set.

For $a, b, \varepsilon \in \mathbb{R}_+$, where $\varepsilon > 0$, let $B_{a,b,\varepsilon} = \{(x,y) \mid a \leq x < a+\varepsilon \wedge b \leq y < b+\varepsilon\}$.

For a given $\varepsilon > 0$, we can decompose the positive quadrant $Q = \{(x,y) \mid x \in \mathbb{R}_+, y \in \mathbb{R}_+\}$ into countably many pairwise disjoint sets as follows

$$Q = \bigcup \mathcal{B}(\varepsilon)\,, \quad \mathcal{B}(\varepsilon) = \{B_{\varepsilon m, \varepsilon n, \varepsilon} \mid m, n \in \mathbb{N}\}\,.$$

For $B \in \mathcal{B}(\varepsilon)$, define

$$\mathcal{X}(B) = \left\{x \in \mathcal{X}^+ \mid P(x) \in B\right\}\,,$$

$$p_I(B) = \int\limits_{\mathcal{X}(B)} p_I(x) dx\,.$$

In analogy to $\mathcal{X}^+, \mathcal{X}_0, \mathcal{X}_1, \mathcal{X}_2$, define

$$\begin{aligned}\mathcal{B}^+(\varepsilon) &= \{B \in \mathcal{B}(\varepsilon) \mid p_I(B) > 0\}\,,\\ c(B) &= \frac{1}{p_I(B)} \int\limits_{\mathcal{X}(B)} p_I(x) c(x) dx \quad \forall B \in \mathcal{B}^+(\varepsilon)\,,\\ \mathcal{B}_0(\varepsilon) &= \{B \in \mathcal{B}^+(\varepsilon) \mid c(B) = 0\}\,,\\ \mathcal{B}_1(\varepsilon) &= \{B \in \mathcal{B}^+(\varepsilon) \mid c(B) = 1\}\,,\\ \mathcal{B}_2(\varepsilon) &= \{B \in \mathcal{B}^+(\varepsilon) \mid 0 < c(B) < 1\}\,.\end{aligned}$$

The set $\mathcal{B}^+(\varepsilon)$ can thus be viewed as a discretisation of $\mathcal{X}$.

Since $a \cdot p_I(x) \le p_O(x) \le (a+\varepsilon) \cdot p_I(x)$ and $b \cdot p_I(x) \le R(x) \le (b+\varepsilon) \cdot p_I(x)$ for all $x \in \mathcal{X}(B_{a,b,\varepsilon})$, it holds

$$a \cdot p_I(B_{a,b,\varepsilon}) c(B_{a,b,\varepsilon}) \le \int\limits_{\mathcal{X}(B_{a,b,\varepsilon})} p_O(x) c(x) dx \le (a+\varepsilon) \cdot p_I(B_{a,b,\varepsilon}) c(B_{a,b,\varepsilon})\,, \tag{19}$$

$$b \cdot p_I(B_{a,b,\varepsilon}) c(B_{a,b,\varepsilon}) \le \int\limits_{\mathcal{X}(B_{a,b,\varepsilon})} R(x) c(x) dx \le (b+\varepsilon) \cdot p_I(B_{a,b,\varepsilon}) c(B_{a,b,\varepsilon})\,. \tag{20}$$

Define

$$\check{P}(B_{a,b,\varepsilon}) = (a, b)\,,$$
$$\hat{P}(B_{a,b,\varepsilon}) = (a+\varepsilon, b+\varepsilon)\,,$$

i.e., $\check{P}(B_{a,b,\varepsilon})$ and $\hat{P}(B_{a,b,\varepsilon})$ is the bottom-left and top-right corner of $B_{a,b,\varepsilon}$, respectively.

Claims 6.1 and 6.2 can be generalized to elements of $\mathcal{B}^+(\varepsilon)$ as follows.

Claim 6.3. *Let $B_{a,b,\varepsilon}, B_{a',b',\varepsilon} \in \mathcal{B}^+(\varepsilon)$, $a \ge a' + \varepsilon$, and $b > b' + \varepsilon$. Then, $B_{a,b,\varepsilon} \in \mathcal{B}_0(\varepsilon)$ or $B_{a',b',\varepsilon} \in \mathcal{B}_1(\varepsilon)$.*

Proof of the claim. Denote $B = B_{a,b,\varepsilon}$, $B' = B_{a',b',\varepsilon}$. By contradiction. Assume $c(B) > 0$ and $c(B') < 1$. Find a selective function c' which is identical to c up to $c'(B) = c(B) - \Delta$, $c'(B') = c(B) + \frac{p_I(B)}{p_I(B')}\Delta$, where

$$\Delta = \min\left\{ c(B), \frac{p_I(B')}{p_I(B)}(1 - c(B')) \right\} > 0\,.$$

Observe that

$$\phi(c') - \phi(c) = -\Delta \cdot p_I(B) + \frac{p_I(B)}{p_I(B')}\Delta \cdot p_I(B') = 0\,.$$

With the use of (19) and (20), derive

$$\begin{aligned}\rho(c') - \rho(c) &= \int\limits_{\mathcal{X}(B')} p_O(x) c'(x) dx - \int\limits_{\mathcal{X}(B)} p_O(x) c(x) dx \le (a'+\varepsilon) p_I(B') c(B') - a \cdot p_I(B) c(B) \\ &\le a(\phi(c') - \phi(c)) = 0\,, \quad \text{and}\end{aligned}$$

$$\begin{aligned}\phi(c)\left(\mathrm{R^S}(h, c') - \mathrm{R^S}(h, c)\right) &= \int\limits_{\mathcal{X}(B')} R(x) c'(x) dx - \int\limits_{\mathcal{X}(B)} R(x) c(x) dx \\ &\le (b'+\varepsilon) p_I(B') c(B') - b \cdot p_I(B) c(B) \\ &< b \cdot (\phi(c') - \phi(c)) = 0\,.\end{aligned}$$

Hence, (h, c') contradicts the optimality of (h, c). ∎

Claim 6.4. *Let $\varepsilon > 0$ and B_1, B_2, B_3 be elements of $\mathcal{B}^+(\varepsilon)$.*

- *If $\beta \cdot \check{P}(B_3) = \alpha_1 \cdot \hat{P}(B_1) + \alpha_2 \cdot \hat{P}(B_2)$, where $\alpha_1, \alpha_2, \beta \in \mathbb{R}_+$, $\alpha_1 + \alpha_2 = 1$, $\beta < 1$, then $B_3 \in \mathcal{B}_0(\varepsilon)$ or $\{B_1, B_2\} \cap \mathcal{B}_1(\varepsilon) \ne \emptyset$.*
- *If $\beta \cdot \hat{P}(B_3) = \alpha_1 \cdot \check{P}(B_1) + \alpha_2 \cdot \check{P}(B_2)$, where $\alpha_1, \alpha_2, \beta \in \mathbb{R}_+$, $\alpha_1 + \alpha_2 = 1$, $\beta > 1$, then $B_3 \in \mathcal{B}_1(\varepsilon)$ or $\{B_1, B_2\} \cap \mathcal{B}_0(\varepsilon) \ne \emptyset$.*

Proof of the claim. Apply the technique from the proof of Claim 6.3 to the proof of Claim 6.2. ∎

For $\varepsilon > 0$, define

$$C_0(\varepsilon) = \bigcup \mathcal{B}_0(\varepsilon)\,, \quad C_1(\varepsilon) = \bigcup \mathcal{B}_1(\varepsilon)\,, \quad C_2(\varepsilon) = \bigcup \mathcal{B}_2(\varepsilon)\,.$$

For $\varepsilon > \varepsilon' > 0$, let $B \in \mathcal{B}^+(\varepsilon)$, $B' \in \mathcal{B}^+(\varepsilon')$, $B' \subset B$. Observe that $B \in \mathcal{B}_0(\varepsilon)$ and $p_I(B') > 0$ implies $B' \in \mathcal{B}_0(\varepsilon')$. And similarly, $B \in \mathcal{B}_1(\varepsilon)$ and $p_I(B') > 0$ implies $B' \in \mathcal{B}_1(\varepsilon')$. This means that

$$\begin{aligned} C_2\left(\frac{\varepsilon}{2}\right) &\subseteq C_2(\varepsilon)\,, \\ C_0\left(\frac{\varepsilon}{2}\right) \cup C_2\left(\frac{\varepsilon}{2}\right) &\subseteq C_0(\varepsilon) \cup C_2(\varepsilon)\,, \\ C_1\left(\frac{\varepsilon}{2}\right) \cup C_2\left(\frac{\varepsilon}{2}\right) &\subseteq C_1(\varepsilon) \cup C_2(\varepsilon)\,. \end{aligned}$$

We can thus define

$$\begin{aligned} C_2 &= \lim_{n\to\infty} C_2\left(\frac{\varepsilon}{2^n}\right)\,, \\ C_0 &= \left(\lim_{n\to\infty}\left[C_0\left(\frac{\varepsilon}{2^n}\right) \cup C_2\left(\frac{\varepsilon}{2^n}\right)\right]\right) \setminus C_2\,, \\ C_1 &= \left(\lim_{n\to\infty}\left[C_1\left(\frac{\varepsilon}{2^n}\right) \cup C_2\left(\frac{\varepsilon}{2^n}\right)\right]\right) \setminus C_2\,. \end{aligned}$$

where we utilize the fact: if a sequence of sets $\{D_n\}_{n=0}^{\infty}$ fulfills $D_{n+1} \subseteq D_n \subseteq \mathbb{R}^2$ for all $n \in \mathbb{N}$, then $\lim_{n\to\infty} D_n = \bigcap_{n=0}^{\infty} D_n$.

Note that each C_i corresponds to A_i (see 15– 17) in the following sense:

$$\int_{\mathcal{X}(A_2 \Delta C_2)} p(x)dx = 0\,, \ \int_{\mathcal{X}((A_0 \cup A_2)\Delta(C_0 \cup C_2))} p(x)dx = 0\,, \ \int_{\mathcal{X}((A_1 \cup A_2)\Delta(C_1 \cup C_2))} p(x)dx = 0\,,$$

where Δ denotes the symmetric difference of two sets.

It holds

$$\dim \operatorname{span}\left((\operatorname{conv}(C_0) \cap \operatorname{conv}(C_1)) \cup C_2\right) < 2\,,$$

otherwise we can find $\varepsilon > 0$ and a configuration of two or three elements of $\mathcal{B}^+(\varepsilon)$ which is ruled out by Claims 6.3 and 6.4 (the analysis of infeasible configurations is analogous to cases 1– 4). $\square$

A.4 Characterization of function τ in Theorem 3

Theorem 7. *Let there be real numbers μ, λ such that (8) fulfills $R(x) + \mu p_O(x) = \lambda p_I(x)$ for all $x \in \mathcal{X}$. Then, there are real numbers $\gamma_1, \gamma_2, \chi_1, \chi_2$, where $\gamma_1 \leq \gamma_2$, and $\chi_1, \chi_2 \in [0, 1]$, such that the selective function τ defined as*

$$\tau(x) = \begin{cases} 1 & \textit{if} \quad \gamma_1 < \frac{p_O(x)}{p_I(x)} < \gamma_2 \\ \chi_1 & \textit{if} \quad \frac{p_O(x)}{p_I(x)} = \gamma_1 \\ \chi_2 & \textit{if} \quad \frac{p_O(x)}{p_I(x)} = \gamma_2 \\ 0 & \textit{otherwise} \end{cases}$$

is an optimal solution to (8).

Proof. Since, for all $x \in \mathcal{X}$, $R(x) = \lambda p_I(x) - \mu p_O(x)$, we can write

$$\mathrm{R^S}(h, c) = \frac{\int_{\mathcal{X}} R(x)c(x)dx}{\phi(c)} = \frac{\lambda\phi(c) - \mu\rho(c)}{\phi(c)} = \lambda - \mu\frac{\rho(c)}{\phi(c)}\,.$$

For $a, b \in \mathbb{R}_+$, let $M_{a,b} = \{x \in \mathcal{X} \mid a < \frac{p_O(x)}{p_I(x)} < b\}$, and $M_a = \{x \in \mathcal{X} \mid \frac{p_O(x)}{p_I(x)} = a\}$. Define continuous functions $\Phi, \mathrm{P} : [0, 1]^2 \times \mathbb{R}_+^2 \to [0, 1]$ as

$$\begin{aligned} \Phi(\alpha, \beta, s, t) &= \alpha \int_{M_s} p_I(x)dx + \int_{M_{s,t}} p_I(x)dx + [\![s < t]\!]\beta \int_{M_t} p_I(x)dx\,, \\ \mathrm{P}(\alpha, \beta, s, t) &= \alpha \int_{M_s} p_O(x)dx + \int_{M_{s,t}} p_O(x)dx + [\![s < t]\!]\beta \int_{M_t} p_O(x)dx\,. \end{aligned}$$

Distinguish two cases.

Case $\mu < 0$. The problem reduces to

$$\min_{h,c} \frac{\rho(c)}{\phi(c)} \qquad \text{s.t.} \qquad \phi(c) \geq \phi_{\min} \quad \text{and} \quad \rho(c) \leq \rho_{\max}\,.$$

An optimal solution τ is obtained by setting

$$\begin{aligned}
\gamma_1 &= 0\,,\\
\gamma_2 &= \inf\left\{t \in \mathbb{R}_+ \mid \Phi(1,1,0,t) \geq \phi_{\min}\right\}\,,\\
\chi_2 &= \begin{cases} \inf\{\beta \in [0,1] \mid \Phi(1,\beta,0,\gamma_2) \geq \phi_{\min}\} & \text{if } \gamma_2 > 0\\ \inf\{\beta \in [0,1] \mid \Phi(\beta,0,0,0) \geq \phi_{\min}\} & \text{otherwise}\end{cases}\,,\\
\chi_1 &= \begin{cases} 1 & \text{if } \gamma_2 > 0\\ \chi_2 & \text{otherwise}\end{cases}\,.
\end{aligned}$$

Note that $\mathrm{P}(\chi_1,\chi_2,\gamma_1,\gamma_2) > \rho_{\max}$ means that the problem is not feasible.

Case $\mu > 0$. The problem reduces to

$$\max_{h,c} \frac{\rho(c)}{\phi(c)} \qquad \text{s.t.} \qquad \phi(c) \geq \phi_{\min} \quad \text{and} \quad \rho(c) \leq \rho_{\max}\,.$$

Define a partial function $F : [0,1] \times \mathbb{R}_+ \to [0,1] \times \mathbb{R}_+$ such that $F(\alpha,s) = (\beta,t)$ iff

$$\begin{aligned}
\mathrm{P}(\alpha,\beta,s,t) &= \rho_{\max}\,,\\
t &= \sup\{a \in \mathbb{R}_+ \mid \mathrm{P}(\alpha,0,s,a) \leq \rho_{\max}\}\,,\\
\beta &= \sup\{b \in [0,1] \mid \mathrm{P}(\alpha,b,s,t) \leq \rho_{\max}\}\,.
\end{aligned}$$

By the assumption that the problem is feasible, an optimal solution τ is obtained by setting

$$\begin{aligned}
\gamma_1 &= \sup\{s \in \mathbb{R}_+ \mid \exists \alpha \in [0,1] : F(\alpha,s) = (\beta,t) \wedge \Phi(\alpha,\beta,s,t) \geq \phi_{\min}\}\,,\\
\chi_1 &= \sup\{\alpha \in [0,1] \mid F(\alpha,\gamma_1) = (\beta,t) \wedge \Phi(\alpha,\beta,\gamma_1,t) \geq \phi_{\min}\}\,,\\
(\chi_2,\gamma_2) &= F(\chi_1,\gamma_1)\,.
\end{aligned}$$

□

A.5 Linear programming formulation of the Bounded TPR-FPR model for finite input sets

Lemma 1. *For any (h,c) optimal to (8), $\phi(c) = \phi_{\min}$ unless $\mathrm{R}^{\mathrm{S}}(h,c) = 0$.*

Proof. By contradiction. Assume that $\mathrm{R}^{\mathrm{S}}(h,c) > 0$ and $\phi(c) = \alpha \cdot \phi_{\min}$ for some $\alpha > 1$. Let c' be the selective function defined by $c'(x) = c(x)/\alpha$ for all $x \in \mathcal{X}$. Then,

$$\begin{aligned}
\phi(c') &= \phi_{\min}\,,\\
\rho(c') &= \frac{\rho(c)}{\alpha} \leq \rho_{\max}\,,\\
\mathrm{R}^{\mathrm{S}}(h,c') &= \frac{\mathrm{R}^{\mathrm{S}}(h,c)}{\alpha} < \mathrm{R}^{\mathrm{S}}(h,c)\,,
\end{aligned}$$

and thus (h,c') contradicts the optimality of (h,c). □

If $\mathcal{X}$ is a finite set, Lemma 1 enables us to refomulate Problem 1 as the following linear program:

$$\min_{c \in [0,1]^{\mathcal{X}}} \sum_{x \in \mathcal{X}} \frac{1}{\phi_{\min}} R(h,x)c(x) \quad \text{s.t.} \quad \sum_{x \in \mathcal{X}} p_I(x)c(x) = \phi_{\min}\,, \quad \sum_{x \in \mathcal{X}} p_O(x)c(x) \leq \rho_{\max}\,. \tag{21}$$

A.6 Proof of Theorem 5

Let (h,c^*) be optimal to (10). Denote $C = \phi(c^*)$. By rewriting (10), it turns out that (h,c^*) is optimal to

$$\min_{h,c} \int_{\mathcal{X}} \frac{1}{C} R(x)c(x)dx \quad \text{s.t.} \quad \begin{array}{lcl} \phi(c) & = & C\\ \rho(c) & \leq & \frac{(1-\pi)(1-\kappa_{\min})}{\pi\kappa_{\min}} C\,. \end{array}$$

According to Lemma 1, this is synonymous with (8), and as a result, Theorem 3 is applicable to c^*.

B Post-hoc tuning and evaluation metrics

B.1 Figures

In Sec. 3, we proposed new evaluation metrics and applied them to synthetic data and exemplar single-score and double-score methods. Synthetic data and exemplar methods are described in Sec 3.1. The proposed evaluation metrics follow from the definition of an optimal OOD selective classifier. We provided two such formulations that avoid the definition of the loss function. Namely, we propose the bounded TPR-FPR rejection model and the Bounded Precision-Recall rejection model.

In case of the bounded TPR-FPR model, the objective, and also the evaluation metric, is the selective risk attained $\mathrm{R}^{\mathrm{S}}_{\mathrm{n}}$ at minimal acceptable TPR $\rho_{\min}$ and maximal acceptable FPR $\rho_{\max}$. In addition to reporting a selective risk for a single operating point, it can be useful to fix the maximal acceptable FPR $\rho_{\max}$ and show the selective risk $\mathrm{R}^{\mathrm{S}}_{\mathrm{n}}$ as the function of varying TPR/coverage $\phi_{\max}$, which yields the Risk-Coverage curve at FPR $\rho_{\max}$. The RC curve at $\rho_{\max} = 0.2$ for our example on synthetic data is shown in Figure 1(a). The proposed double score method $D(\mathbb{R})$ is seen to achieve the lowest selective risk in the entire range of coverages available. The selective risk of the methods $D(\mathbb{R})$ and $C(0)$ is the same; however, the method $C(0)$ has much lower maximal attainable coverage, namely, $\phi_{\max} = 0.58$ and hence the method is marked as unable to achieve the target coverage; see Table 1.

The problem of defining the TPR-FPR model (13) can be infeasible. To choose a feasible target value of $\phi_{\min}$ and $\rho_{\max}$, it is advantageous to plot the ROC curve, that is, the TPR and FPR values attainable by the classifiers in $\mathcal{Q}$. ROC curve for the methods in our example is shown in Figure 1(b). The operation point $(\phi_{\min}, \rho_{\max})$ is attainable if the ROC curve of the given method is entirely above the point.

In case of the bounded Precision-Recall model, the objective, and also the evaluation metric, is the selective risk attained $\mathrm{R}^{\mathrm{S}}_{\mathrm{n}}$ at minimal acceptable Precision $\kappa_{\min}$ and minimal acceptable Recall/TPR $\phi_{\min}$. In our example, the single-score method achieves the same selective risk under both models as we use the same target TPR/recall and the selective risk is a monotonic function of the score, see discussion in sec 3.3, hence we do not show the risk-coverage curve at fixed precision. However, we show the Precision-Recall curve, Figure 1(c), which is useful for determining the feasible target value for precision and recall. Again, the operation point $(\kappa_{\min}, \phi_{\min})$ is achievable if the PR curve of the given method is entirely above the point.

B.2 Algorithms

The single-score OODD methods output a set of selective OOD classifiers $\mathcal{Q} = \{(h, c) \mid c(x) = [\![s(x) \leq \lambda]\!], \lambda \in \mathbb{R}\}$ parameterized by the decision threshold λ. Double-score OODD methods output a set $\mathcal{Q} = \{(h, c) \mid c(x) = [\![s_r(x) + \mu\, s_g(x) \leq \lambda]\!], \mu \in \mathbb{R}, \lambda \in \mathbb{R}\}$ parameterized by $\lambda \in \mathbb{R}$ and $\mu \in \mathbb{R}$.

The post hoc tuning aims to find the best OOD selective classifier out of $\mathcal{Q}$ based on the appropriate metric. To this end, the existing methods used the AUROC, AUPR of OSCR score as the metric to find the best classifier. Instead, we formulate the bounded TPR-FPR and the bound Precision-Recall model, where we find the best selective amounts to solving the constrained optimization problem (13) and (14), respectively.

In case of single-score methods, the problem (13) and (14) 1-D optimization, namely, one needs to find the decision threshold $\lambda \in \mathbb{R}$ which leads to the minimal selective risk and simultaneously satisfies both constraints on the validation set $\mathcal{T} = ((x_i, \bar{y}_i) \in \mathcal{X} \times \bar{\mathcal{Y}} \mid i = 1, \dots, n)$. The threshold λ influences the involved metrics, that is, ($\mathrm{R}^{\mathrm{S}}_{\mathrm{n}}(h, c)$, $\phi_n(c)$, $\rho_n(c)$, $\kappa_n(c)$), only via the value of the selective function $c(x) = [\![s(x) \leq \lambda]\!]$ which is a step function of the optimized threshold λ. Hence, we can see ($\mathrm{R}^{\mathrm{S}}_{\mathrm{n}}(\lambda)$, $\phi_n(\lambda)$, $\rho_n(\lambda)$, $\kappa_n(\lambda)$), as a function of λ and we can find all $n+1$ achievable values of ($\mathrm{R}^{\mathrm{S}}_{\mathrm{n}}(\lambda)$, $\phi_n(\lambda)$, $\rho_n(\lambda)$, $\kappa_n(\lambda)$) in a single sweep over the validation examples $\mathcal{T}$ sorted according to the value of $s(x_i)$. This procedure has complexity $\mathcal{O}(n \log n)$ attributed to the sorting of n examples.

In case of the double-score methods, we need to optimize w.r.t. λ and μ which are the free parameters of the selective function $c(x) = [\![s_r(x) + \mu\, s_g(x) \leq \lambda]\!]$. The selective classifier can be seen as a binary in 2-D space. Hence, we equivalently parameterize the selective function as $c(x) =$

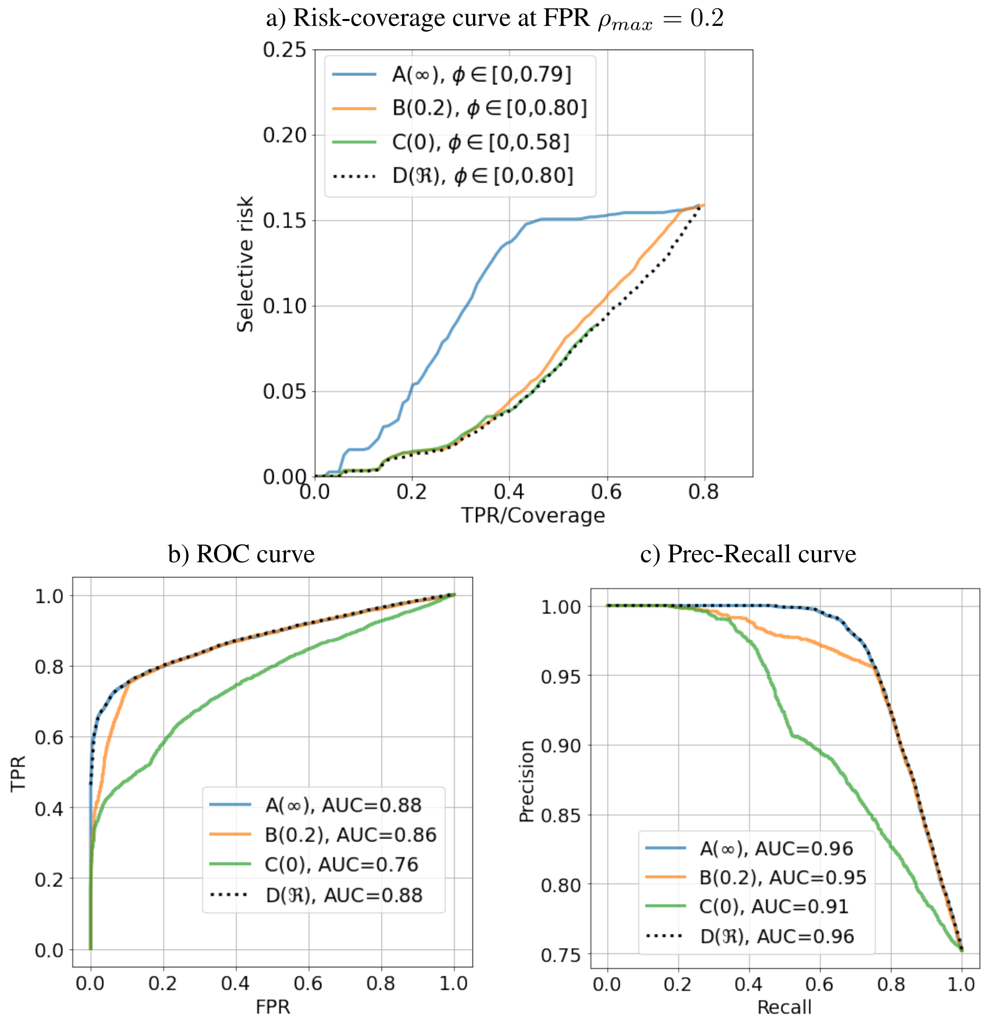

Figure 1: Evaluation curves for the exemplar methods on the synthetic data. The risk-coverage/TPR curve at the maximal acceptable FPR is shown in Fig. a). For each method we also show attainable coverage ϕ. The ROC curve and the Precision-Recall curve are shown in Fig. b) and c), respectively.

$[\![s_r(x), \cos(\alpha) + s_g(x) \sin(\alpha) \leq \lambda']\!]$ where $\alpha \in \mathcal{A} = [0, \pi]$ and $\lambda' \in \mathbb{R}$. We approximate $\mathcal{A}$ by a finite set $\bar{\mathcal{A}} \subset \mathcal{A}$, where $\bar{\mathcal{A}}$ contains d equidistantly placed values over the interval $[0, \pi]$. For each $\alpha \in \bar{\mathcal{A}}$, we compute all values $n + 1$ of($\mathrm{R}^{\mathrm{S}}_{\mathrm{n}}(\lambda)$, $\phi_n(\lambda)$, $\rho_n(\lambda)$, $\kappa_n(\lambda)$), using the algorithm described above. We found that setting $d = 360$ is enough, as higher values d do not change the results.