Appendix

A. Proofs

A.1. Proofs for Section 2

Proof of Theorem 1. Consider the distribution where there are two connected components, one for $\mathcal{X}_+$ and one for $\mathcal{X}_-$, each with mixture probability $p = \frac{1}{2}$. Thus, Assumption 1 holds and we are choose to free the other parameters of the distribution in any way that satisfies Assumption 2 and 3 (e.g. a mixture of uniform density functions satisfies these assumptions). Now note that with probability $\frac{1}{2}$, the final point that is label queried by the passive learner will be positive and, thus, the passive algorithm will need to query all of the points with probability $\frac{1}{2}$ in order to retrieve all positive points. In such an event, the excess query cost is at least $\frac{1}{2} \cdot n$. □

A.2. Proofs for Section 3

Much of our technical results require the following uniform high-probability guarantee that balls of sufficient probability mass contain an example:

Lemma 3. *Let $0 < \delta < 1$ and $\mathcal{F}$ be some distribution over $\mathbb{R}^D$ and X be a sample of size n drawn i.i.d. from $\mathcal{F}$. There exists universal constant C_0 such that the following holds with probability at least $1 - \delta$ uniformly over all balls $B \in \mathbb{R}^D$:*

$$\mathcal{F}(B) \geq \frac{C_0 \cdot D \cdot \log(2/\delta) \log n}{n} \Rightarrow |B \cap X| > 0.$$

Proof. This follows by Lemma 7 of Chaudhuri & Dasgupta (2010). □

The following result bounds the volume of the ϵ-neighborhood around $\mathcal{X}_+$, which will be used later to bound the excess number of points queried around $\mathcal{X}_+$. The result says that the volume of the ϵ-neighboorhood around $\mathcal{X}_+$ (and not including $\mathcal{X}_+$) is linear in ϵ.

Lemma 4. *Suppose Assumption 2 holds. Then there exists constants $r_1, C'_+ > 0$ depending only on $\mathcal{X}_+$ such that for all $0 < \epsilon < r_1$, we have*

$$\textit{Vol}(B(\mathcal{X}_+, \epsilon) \backslash \mathcal{X}_+) \leq C'_+ \cdot \epsilon,$$

where $B(\mathcal{X}_+, \epsilon) := \{x \in \mathbb{R}^D : \inf_{x' \in \mathcal{X}_+} |x - x'| \leq \epsilon\}$.

Proof of Lemma 4. This follows from Gorin (1983). To see this, the equation on page 159 of Gorin (1983) states that if M and N are respectively d-dimensional and $(d + k)$-dimensional compact smooth Riemannian manifolds and $f : M \to N$ is a smooth isometric embedding, then we have

$$\text{Vol}(B(f(M), \epsilon)) = V_k \cdot \epsilon^k \cdot \text{Vol}(M) + O(\epsilon^{k+1}),$$

where V_k is the volume of a k-dimensional ball. Here, we take $M = \mathcal{X}_+$ and $N = B(\mathcal{X}_+, r_1)$ for some $r_1 > 0$. Then, we have $k = 0$ and taking f to be the identity function, we have

$$\text{Vol}(B(\mathcal{X}_+, \epsilon)) = \text{Vol}(\mathcal{X}_+) + O(\epsilon),$$

and the result follows immediately. □

Proof of Theorem 2. By Hoeffding's inequality, out of the initial m examples that Algorithm 1 label queries, we have with probability at least $1 - \delta/2$ that at least $p - \sqrt{\frac{1}{2m} \cdot \log(2/\delta)}$ fraction of them are positively labeled, since the example being positive follows a Bernoulli distribution with probability p. Then by the condition on m, we have that at least $p/2$ fraction of the points are positively labeled.

Take

$$\epsilon = \left(\frac{2 \cdot C_0 \cdot D \cdot \log(4/\delta) \log(p \cdot m/2)}{p^2 \cdot \lambda_0 \cdot C_+ \cdot v_D \cdot m} \right)^{1/D}, \qquad M_0 = \max\left\{ \frac{2 \cdot C_0 \cdot D(\log(p \cdot /2) + 1)}{p^2 \cdot \lambda_0 \cdot C_+ \cdot v_D \cdot \min\{r_0, r_1\}^D}, 2e \right\},$$

where v_D is the volume of a unit ball in $\mathbb{R}^D$. Then, we have that the condition on m and M_0 guarantees that $\epsilon < \min\{r_0, r_1\}$.

Let $x \in \mathcal{X}_+$. We have that the probability mass of positive examples in $B(x, \epsilon)$ w.r.t. $\mathcal{P}$ is:

$$\begin{aligned} p \cdot \mathcal{P}_+(B(x,\epsilon)) &\geq p \cdot \lambda_0 \cdot \text{Vol}(B(x,\epsilon) \cap \mathcal{X}_+) \\ &\geq p \cdot \lambda_0 \cdot C_+ \cdot \text{Vol}(B(x,\epsilon)) \\ &\geq p \cdot \lambda_0 \cdot C_+ \cdot v_D \cdot \epsilon^D \\ &\geq \frac{2 \cdot C_0 \cdot D \log(4/\delta) \log(p \cdot m/2)}{p \cdot m}. \end{aligned}$$

Then by Lemma 3, we have with probability at least $1 - \delta/2$ that all the positive examples in X are within ϵ of one of the positive examples among the initially sampled m examples. Therefore, Algorithm 1 retrieves all of the positive examples.

Now we bound the expected regret:

$$\begin{aligned} \mathbb{E}[C_{\text{offline}}] &\leq (1-p) \cdot m + n \cdot (1-p) \cdot \mathcal{P}_-(B(\mathcal{X}_+, \epsilon) \backslash \mathcal{X}_+) \\ &\leq (1-p) \cdot m + n \cdot (1-p) \cdot \lambda_1 \cdot C'_+ \cdot \epsilon. \end{aligned}$$

The result follows. □

Proof of Theorem 3. Let $\mathcal{P}_+$ be the uniform distribution on the unit hypercube $[0,1]^D$ and $\mathcal{P}_-$ be the uniform distribution on $[-1,2]^D$. In the initial sampling phase of Algorithm 1, at most m of the examples will be positively labeled. Let $\widehat{\mathcal{X}_+} = X \cap \left(\cup_{x \in X_{0,+}} B(x,\epsilon)\right)$, the set of points that Algorithm 1 labeled. Then, Theorem 3b in (Cuevas et al., 1997) shows that for n sufficiently large, with probability at least $1/4$, we have

$$d_H(\widehat{\mathcal{X}_+}, \mathcal{X}_+) \geq \frac{1}{4}\left(\frac{\log m}{m}\right)^{1/D}$$

for any $\epsilon > 0$, where $d_H(A, B) := \max\{\sup_{x\in A} d(x, B), \sup_{x \in B} d(x, A)\}$ is the Hausdorff distance. Therefore, we have (in the case of taking $\epsilon \to 0$):

$$d_H(X_{0,+}, \mathcal{X}_+) \geq \frac{1}{4}\left(\frac{\log m}{m}\right)^{1/D}.$$

Since $X_{0,+} \subseteq \mathcal{X}_+$, it follows that $d_H(X_{0,+}, \mathcal{X}_+) = \sup_{x \in \mathcal{X}_+} d(x, X_{0,+})$. Therefore, we need $\epsilon \geq \frac{1}{4}\left(\frac{\log m}{m}\right)^{1/D}$ in order for Algorithm 1 to recover all of the positive examples. Thus, the expected regret is at least (for some $C > 0$)

$$\mathbb{E}[C_{\text{offline}}] \geq (1-p) \cdot m + C \cdot \left(\frac{\log m}{m}\right)^{1/D} \cdot n,$$

as desired. □

A.3. Proofs for Section 4

Proof of Lemma 1. By Hoeffding's inequality, out of the initial m examples that Algorithm 2 label queries, we have with probability at least $1 - \delta/2$ that at least $p - \sqrt{\frac{1}{2m} \cdot \log(2/\delta)}$ are fraction of them are positively labeled, since the example being positive follows a Bernoulli distribution with probability p. Then by the condition on m, we have that at least $p/2$ fraction of the points are positively labeled and thus we have at least $mp/2$ positive examples.

Then, we have that out of these $mp/2$ examples, the probability that none of them are in $\mathcal{X}_{+,i}$ for each $i \in [c]$ is at most

$$(1 - \mathcal{P}_+(\mathcal{X}_{+,i}))^{mp/2} \leq (1-q)^{mp/2} \leq \frac{\delta}{2c}.$$

The result follows by union bound. □

Proof of Lemma 2. Let $x, x' \in \mathcal{X}_{+,i}$. There exists a path $x = x_1 \to x_2 \to \ldots \to x_q = x'$ in $\mathcal{X}_{+,i}$ such that $||x_j - x_{j+1}|| \leq \epsilon/3$. We also have that the probability mass of positive examples in $B(x_j, \epsilon/3)$ w.r.t. $\mathcal{P}$ is:

$$p \cdot \mathcal{P}_+(B(x_j, \epsilon/2)) \geq p \cdot C_+ \cdot \lambda_0 \cdot v_D \cdot (\epsilon/3)^D \geq \frac{C_0 \cdot D \log(2/\delta) \cdot \log n}{n}.$$

Therefore by Lemma 3, there exists $x'_j \in B(x_j, \epsilon/3) \cap X_+$, where X_+ are the positive examples in X. Hence, by triangle inequality, there exists path $x = x'_1 \to x'_2 \to ... \to x'_q = x'$ all in X_+ where $||x'_j - x'_{j+1}|| \leq ||x'_j - x_j|| + ||x_j - x_{j+1}|| + ||x_{j+1} - x'_{j+1}|| \leq \epsilon$ implying that $\mathcal{X}_{+,i} \cap X_+$ is connected in the ϵ-neighborhood graph of X_+. The result follows immediately. □

Finally, we combine these two results to show the final excess query cost guarantee for Algorithm 2.

Proof of Theorem 4. Take

$$N_0 = \frac{3^D \cdot C_0 \cdot D}{\min\{r_0, r_1\}^D \cdot p \cdot C_+ \cdot \lambda_0 \cdot v_D}.$$

By Lemma 1, there exists at least one positive example in the initial m samples from each connected component of $\mathcal{X}_+$. Define

$$\epsilon := 3 \left(\frac{C_0 \cdot D \log(4/\delta) \cdot \log n}{p \cdot C_+ \cdot \lambda_0 \cdot v_D \cdot n} \right)^{1/D}.$$

We have that the condition on n implies that $\epsilon \leq \min\{r_0, r_1\}$. By Lemma 2, we have that all of the positive examples of each connected component of $\mathcal{X}_+$ are in the same CC of the ϵ-neighborhood graph of the positive examples. Therefore, when the algorithm terminates, the set of examples it will select from will be contained in $B(\mathcal{X}_+, \epsilon)$.

Therefore, we have

$$\mathbb{E}[C_{\text{exp-commit}}] \leq (1-p) \cdot m + C'_+ \cdot \epsilon \cdot n \leq (1-p) \cdot m + C \cdot ((\log(4/\delta) \cdot \log n)^{1/D} \cdot n^{(D-1)/D},$$

for some C depending on $\mathcal{P}$, as desired. □

B. Additional Experiment Plots

In Table 3, we show the full results for the Letters dataset for the area under curve metrics. We see that in all cases, our method outperforms outright. In Table 4, we show the full results for CelebA. We see that our method is competitive for 32 out of the 40 tasks.

Dataset	Label	O	O-LS	A-LS	O-RS	A-RS	O-IF	A-IF	O-RC	A-RC	EC (Ours)
Letters	A	91.14	52.52	52.49	84.81	89.02	59.52	84.69	64.27	86.87	**97.12**
	B	83.41	52.73	52.57	75.95	82.41	56.13	75.89	61.38	80.18	**93.76**
	C	84.48	56.19	56.08	75.92	83.78	55.78	78.68	59.21	81.92	**94.2**
	D	83.51	52.57	52.45	76.14	82.23	56.09	76.03	61.51	78.83	**93.73**
	E	78.6	52.85	52.99	74.84	81.41	55.77	73.7	60.17	77.27	**89.5**
	F	83.63	53.4	53.41	78.72	83.16	57.44	77.83	64.69	80.88	**94.0**
	G	82.23	52.72	52.67	75.76	81.88	58.15	74.58	63.45	79.79	**92.35**
	H	69.07	52.4	52.63	61.19	69.51	53.95	64.15	52.03	66.94	**81.78**
	I	84.26	52.37	52.59	73.96	80.72	55.95	77.56	59.11	83.83	**93.89**
	J	84.24	54.45	54.35	75.16	81.72	56.22	77.06	58.58	81.61	**94.28**
	K	72.7	52.57	52.62	65.15	72.03	55.46	68.12	51.45	74.38	**89.4**
	L	80.98	52.59	52.67	72.74	79.87	55.44	77.73	59.16	78.57	**93.18**
	M	81.83	54.41	54.25	77.48	83.23	59.55	78.61	60.62	81.67	**93.07**
	N	75.15	52.67	52.68	68.58	75.31	55.37	69.43	55.85	73.74	**89.8**
	O	87.71	52.51	52.62	80.88	86.08	57.88	78.08	67.62	82.71	**94.69**
	P	84.76	52.7	52.7	79.24	84.47	57.63	79.18	64.81	83.19	**93.47**
	Q	82.25	53.22	53.08	74.97	80.09	55.71	74.66	60.49	79.55	**92.18**
	R	82.68	52.59	52.48	76.72	82.47	56.32	75.3	61.37	79.26	**92.77**
	S	76.51	52.83	52.91	70.42	75.85	55.24	72.01	59.66	75.84	**89.27**
	T	83.47	55.32	55.35	76.38	82.95	56.87	79.06	62.87	81.96	**93.34**
	U	78.05	52.91	53.05	73.72	81.28	55.86	72.81	58.56	76.41	**92.7**
	V	89.82	54.61	54.59	82.3	87.65	59.74	82.18	61.01	85.18	**96.19**
	W	90.15	55.25	55.29	84.57	89.11	59.37	82.04	63.1	86.65	**96.79**
	X	80.18	52.63	52.6	74.84	80.68	55.92	73.04	62.48	78.1	**92.93**
	Y	81.73	54.65	54.7	72.3	79.67	57.49	76.38	53.48	79.33	**92.46**
	Z	83.78	54.6	54.54	76.89	84.18	56.46	78.56	60.14	82.64	**93.35**

Table 3. **Letters**: Area under curve metric.

Label	O	O-LS	A-LS	O-RS	A-RS	O-IF	A-IF	O-RC	A-RC	EC (Ours)
5-o-Clock-Shadow	20.17	15.54	15.55	17.31	18.28	17.89	18.85	20.25	22.49	**23.85**
Arched-Eyebrows	19.29	15.52	15.51	17.54	17.69	18.34	18.65	19.87	19.76	**20.93**
Attractive	18.4	**19.19**	18.32	18.6	18.7	**18.86**	**18.89**	18.03	17.93	18.61
Bags-Under-Eyes	17.06	15.5	15.51	16.58	**17.54**	**17.26**	17.11	**17.3**	16.33	16.97
Bangs	**20.88**	15.52	15.52	**20.6**	19.97	20.49	**21.43**	**22.12**	19.9	**22.08**
Bald	18.52	15.48	15.49	16.6	16.98	15.48	17.92	NA	NA	**28.34**
Big-Lips	15.81	15.51	15.5	15.44	14.91	15.26	15.46	15.61	15.68	**16.35**
Big-Nose	16.58	15.53	15.53	15.92	16.58	15.95	16.23	16.88	16.12	**17.23**
Black-Hair	22.87	15.49	15.5	20.8	19.43	21.35	22.52	21.23	20.5	**24.6**
Blond-Hair	36.36	15.67	15.67	31.41	34.64	35.92	37.84	37.31	39.24	**41.72**
Blurry	16.75	15.44	15.48	16.38	16.06	16.48	16.66	16.66	16.81	**17.79**
Brown-Hair	20.88	15.48	15.48	18.82	20.45	20.64	**21.34**	20.94	20.77	**21.57**
Bushy-Eyebrows	19.11	15.49	15.49	17.87	18.12	18.18	18.7	18.88	19.2	**21.3**
Chubby	16.77	15.51	15.5	15.98	16.2	15.83	16.15	16.51	**19.01**	**18.63**
Double-Chin	18.05	15.53	15.52	16.73	17.31	16.17	17.28	16.66	**22.57**	**22.3**
Eyeglasses	16.48	15.51	15.5	16.14	**17.12**	**16.48**	**16.78**	**17.74**	15.88	15.44
Goatee	17.57	15.49	15.51	16.86	16.69	16.22	16.93	16.98	17.98	**19.81**
Gray-Hair	23.19	15.53	15.55	19.55	21.77	17.01	23.12	20.84	**32.86**	**31.66**
Heavy-Makeup	21.42	15.85	15.58	20.49	20.5	21.24	**21.94**	21.29	20.4	**22.45**
High-Cheekbones	19.17	15.38	15.17	18.59	18.79	18.91	18.89	19.78	18.77	**20.07**
Male	18.64	15.49	15.57	**20.09**	19.36	19.14	18.8	18.43	16.5	19.24
Mouth-Slightly-Open	17.37	15.66	15.56	17.43	17.34	17.42	17.16	**17.79**	17.2	**17.77**
Mustache	17.16	15.5	15.48	16.77	16.46	15.98	16.99	16.64	17.05	**18.13**
Narrow-Eyes	15.48	15.49	15.5	15.66	**15.95**	**15.78**	**15.78**	15.68	15.36	14.83
No-Beard	15.82	**16.38**	**16.37**	15.79	15.93	16.0	15.99	16.08	15.98	15.83
Oval-Face	18.02	15.5	15.5	16.76	17.11	17.18	17.37	18.56	18.37	**19.22**
Pale-Skin	18.15	15.45	15.48	**19.89**	**19.15**	16.95	**20.95**	**18.75**	17.94	**19.55**
Pointy-Nose	18.55	15.49	15.49	17.03	17.22	17.23	17.75	18.69	18.8	**19.97**
Receding-Hairline	18.5	15.5	15.5	17.09	17.47	16.67	17.55	18.61	**22.2**	**22.38**
Rosy-Cheeks	23.96	15.5	15.51	19.77	22.49	18.5	22.3	22.45	32.64	**34.69**
Sideburns	18.32	15.51	15.52	17.23	17.19	16.9	17.39	17.41	19.36	**21.77**
Smiling	19.46	16.01	15.52	18.97	19.05	19.31	19.15	19.89	18.81	**20.15**
Straight-Hair	**16.1**	15.5	15.5	**15.97**	**16.25**	15.91	**16.02**	**16.24**	**16.13**	**16.29**
Wavy-Hair	21.0	15.52	15.52	20.18	20.09	20.56	20.97	21.13	20.42	**21.88**
Wearing-Earrings	18.75	15.47	15.47	16.93	17.87	17.75	18.37	19.28	20.0	**20.62**
Wearing-Hat	**18.32**	15.48	15.49	**18.33**	**18.27**	17.44	**19.95**	**20.12**	15.16	16.82
Wearing-Lipstick	20.42	19.0	17.34	19.64	19.65	20.49	**20.8**	20.21	19.51	**21.06**
Wearing-Necklace	18.99	15.48	15.49	16.72	17.85	17.54	18.15	19.44	**21.28**	**21.48**
Wearing-Necktie	19.56	15.52	15.51	17.51	18.13	16.53	18.36	18.45	**23.79**	**24.46**
Young	15.51	**16.22**	**16.21**	15.54	15.63	15.55	15.55	15.56	15.51	15.57

Table 4. **CelebA**: Area under curve metric. We note that for Bald, there were no results for the Robust Covariance metrics. This is because due to the low rate of positive examples, it was not possible to tune Robust Covariance's hyper-parameters via cross-validation on the initial sample.