A Proof of Proposition 1

Our proof follows the standard proof of MAB lower bound. Let Δ be some constant. Define some $\theta_1 = (\underbrace{\Delta, \dots, \Delta}_{T \text{ times}}, \underbrace{0, \dots, 0}_{T(T-1) \text{ times}})^\top$. For any algorithm running on this ASD, let $T_1(n)$ be the total number of selections for the first T samples. If $T_1(n) \leq T/2$, the regret for θ_1 is greater than $\Delta T/2$. Let

$$k = \arg\min_{t>1} \mathbb{E}_{\theta_1} T_t(n).$$

Since $\sum_t \mathbb{E}_{\theta_1} T_t(n) = T$, we have $\mathbb{E}_{\theta_1} T_k(n) \leq T/(T-1) \leq 2$. Without a loss of generality, we let $k = 2$. Define $\tilde{\theta}_2 = (\underbrace{0, \dots, 0}_{T \text{ times}}, \underbrace{2\Delta, \dots, 2\Delta}_{T \text{ times}}, \underbrace{0, \dots, 0}_{T(T-2) \text{ times}})^\top$. Then consider a uniform prior over $\{\theta_1, \tilde{\theta}_2\}$. The Bayesian regret is at least

$$\frac{T\Delta}{4} \mathbb{P}_{\theta_1}(T_1(n) \leq T/2) + \frac{T\Delta}{4} \mathbb{P}_{\tilde{\theta}_2}(T_1(n) > T/2)$$

where $\mathbb{P}_\theta$ is the probability measure under the ASD problem with parameter θ. Using Lemma 15.1 [27], we have

$$\begin{aligned}
\mathbb{P}_{\theta_1}(T_1(n) \leq T/2) + \mathbb{P}_{\tilde{\theta}_2}(T_1(n) > T/2) &\geq \exp\left(-\mathrm{D}\left(\mathbb{P}_{\theta_1}, \mathbb{P}_{\tilde{\theta}_2}\right)\right)/2 \\
&= \exp\left(-\mathbb{E}_{\theta_1}\left[T_2(n)\right] \mathrm{D}(\mathcal{N}(0,1), \mathcal{N}(2\Delta, 1))\right)/2 \\
&\geq \exp(-4\Delta^2)/2,
\end{aligned}$$

where $D(P_1, P_2)$ is the KL-divergence of two probability measure P_1 and P_2. Now choosing $\Delta = 1/2$, we have for any algorithm $\mathcal{A}$, $\mathcal{BR}(T, \mathcal{A}) \gtrsim T$.

B Proof of Lemma 1

We decompose the Bayesian regret in terms of the instant regret

$$\begin{aligned}
\mathcal{BR}(T; IDS) &= \mathbb{E}[\sum_{t=1}^{T} f_\theta(X_t^*) - \sum_{t=1}^{T} f_\theta(X_t)] \\
&= \mathbb{E}[\sum_{t=1}^{T} \langle \pi_t, \Delta_t \rangle] \\
&\leq \mathbb{E}[\sum_{t=1}^{T} (\Psi_{\star,\lambda} g_t^\top \pi_t)^{1/\lambda}] \\
&= \Psi_{\star,\lambda}^{1/\lambda} \mathbb{E}[\sum_{t=1}^{T} (g_t^\top \pi_t)^{1/\lambda}] \\
&\leq \Psi_{\star,\lambda}^{1/\lambda} T^{1-1/\lambda} \mathbb{E}[\sum_{t=1}^{T} (g_t^\top \pi_t)]^{1/\lambda} \\
&\leq \Psi_{\star,\lambda}^{1/\lambda} T^{1-1/\lambda} \mathbb{E}[\sum_{t=1}^{T} I_t(X_{t,1}^*, \dots, X_{t,T-t+1}^*; (X_t, Y_t))]^{1/\lambda} \\
&\quad \text{(using the fact that } \{X_{t,1}^*, \dots, X_{t,T-t+1}^*\} \subset \{X_1^*, \dots, X_T^*\}) \\
&\leq \Psi_{\star,\lambda}^{1/\lambda} T^{1-1/\lambda} \mathbb{E}[\sum_{t=1}^{T} I_t(X_1^*, \dots, X_T^*; (X_t, Y_t))]^{1/\lambda} \\
&\leq \Psi_{\star,\lambda}^{1/\lambda} T^{1-1/\lambda} H(X_1^*, \dots X_T^*)^{1/\lambda}
\end{aligned}$$

C Generalized Linear Model

We start from proving for the simple linear regression i.e. $\mu(x) = x$.

Proof. Let π_t^{ts} be the Thompson sample policy at the step t, i.e. $\pi_t^{ts}(x) = \mathbb{P}_t(X_{t,1}^* = x)$. We have

$$\Psi_t = \frac{(\Delta_t^\top \pi_t)^2}{g_t^\top \pi_t} \leq \frac{(\Delta_t^\top \pi_t^{ts})^2}{g_t^\top \pi_t^{ts}}.$$

We first write the instant regret in the following form:

$$\begin{aligned}\Delta_t^\top \pi_t^{ts} &= \mathbb{E}_t[\theta^\top X_{t,1}^*] - \sum_{x \in S_t^n} \pi_t^{ts}(x) \mathbb{E}_t[\theta^\top x] \\ &\leq \sum_{x \in S_t^n} \mathbb{P}_t(X_{t,1}^* = x) \left(\mathbb{E}_t[\theta^\top x \mid X_1^* = x] - \mathbb{E}_t[\theta^\top x]\right).\end{aligned}$$

Now we deal with the information gain term. Let $Y_{t,x}$ be the observation at step t when selecting input x and $\epsilon_{t,x}$ be the noise generated. We have $Y_{t,x} = \mu(\theta^\top x) + \epsilon_{t,x}$. The mutual information can represented by the Kullback-Leibler divergence between the joint distribution of the two variables and the product of their marginal distribution, i.e.

$$I_t(X_{t,1}^*; Y_{t,x}) = \sum_{x' \in S_t^n} D(\mathbb{P}_t(Y_{t,x} \mid X_{t,1}^* = x') \| \mathbb{P}_t(Y_{t,x})). \tag{3}$$

Lemma 2 (Fact 9 [34])**.** *For any distribution P and Q such that P is absolutely continuous with respect to Q, any random variable $X : \Omega \mapsto \mathcal{X}$ and any $g : \mathcal{X} \mapsto \mathbb{R}$ such that* $\sup g - \inf g \leq 1$*,*

$$\mathbb{E}_P[g(X)] - \mathbb{E}_Q[g(X)] \leq \sqrt{\frac{1}{2} D(P\|Q)},$$

where $\mathbb{E}_P$ and $\mathbb{E}_Q$ denote the expectation operators under P and Q.

By Lemma 2 with $g(x) = x$ and $X = Y_{t,x}$, the information gain can be lower bounded by:

$$g_t^\top \pi_t^{ts} \geq 2 \sum_{x, x' \in S_t^n} \mathbb{P}_t(X_{t,1}^* = x) \mathbb{P}_t(X_{t,1}^* = x') \left(\mathbb{E}_t[Y_{t,x} \mid X_{t,1}^* = x'] - \mathbb{E}_t[Y_{t,x}]\right)^2$$

Let

$$M_{x,x'} = \sqrt{P(X_{t,1}^* = x) P(X_{t,1}^* = x')} (\mathbb{E}[Y_{t,x} \mid X_{t,1}^* = x] - \mathbb{E}[Y_{t,x}]).$$

Then $\Delta_t^\top \pi_t^{ts} = \operatorname{trace}(M)$ and $v_t^\top \pi_t^{ts} = \|M\|_F^2$.

By Fact 2 of [35], we have

$$\Delta_t^\top \pi_t^{ts} / v_t^\top \pi_t^{ts} = \operatorname{trace}(M) / (2\|M\|_F^2) \leq \operatorname{rank}(M) \leq d/2.$$

□

C.1 Proof of Theorem 1

Proof. Using the similar strategy for linear model, we let π_t^{ts} be the Thompson sample policy at the step t.

Using the fact that μ is L_μ-Lipschitz

$$\begin{aligned}\Delta_t^\top \pi_t^{ts} &= \mathbb{E}_t[\mu(\theta^\top X_{t,1}^*)] - \sum_{x \in S_t^n} \pi_t^{ts}(x) \mathbb{E}_t[\mu(\theta^\top x)] \\ &= \sum_{x \in S_t^n} \mathbb{P}_t(X_{t,1}^* = x) \left(\mathbb{E}_t[\mu(\theta^\top X_{t,1}^*)] - \mathbb{E}_t[\mu(\theta^\top x)]\right) \\ &\leq L_\mu \sum_{x \in S_t^n} \mathbb{P}_t(X_{t,1}^* = x) \left(\mathbb{E}_t[(\theta^\top X_{t,1}^*)] - \mathbb{E}_t[(\theta^\top x)]\right) \\ &\leq L_\mu \sum_{x \in S_t^n} \mathbb{P}_t(X_{t,1}^* = x) \left(\mathbb{E}_t[(\theta^\top x) \mid X_{t,1}^* = x] - \mathbb{E}_t[(\theta^\top x)]\right).\end{aligned}$$

The major difference is on the lower bound of the information gain term. We follow a similar strategy in [10]. Let $f(x) = \mathbb{E}[\mu^{-1}(x - \tilde{\epsilon})]$, where $\tilde{\epsilon}$ follows exactly the same distribution of ϵ. Let $\tilde{Y}_{t,x} = f(Y_{t,x}) = \mathbb{E}[\mu^{-1}(Y_{t,x} - \tilde{\epsilon}) \mid Y_{t,x}]$. Then we have

$$\begin{aligned}
\mathbb{E}[\tilde{Y}_{t,x} \mid \theta] &= \mathbb{E}[\mathbb{E}[\mu^{-1}(Y_{t,x} - \tilde{\epsilon}) \mid Y_{t,x}] \mid \theta] \\
&= \mathbb{E}[\mathbb{E}[\mu^{-1}(Y_{t,x} - \epsilon) \mid Y_{t,x}] \mid \theta] \\
&= \mathbb{E}[\mathbb{E}[\mu^{-1}(\theta^\top x) \mid Y_{t,x}] \mid \theta] \\
&= \theta^\top x.
\end{aligned}$$

Since $\tilde{Y}_{t,x}$ is a linear regression outcome, the information gain

$$\begin{aligned}
I(X_1^*; \tilde{Y}_{t,x}) &\geq 2 \sum_{x' \in S_t^n} \mathbb{P}_t(X_{t,1}^* = x')(\mathbb{E}_t[\tilde{Y}_{t,x} \mid X_{t,1}^* = x'] - \mathbb{E}_t[\tilde{Y}_{t,x}])^2 \\
&= 2 \sum_{x' \in S_t^n} \mathbb{P}_t(X_{t,1}^* = x')(\mathbb{E}_t[\theta^\top x \mid X_{t,1}^* = x'] - \mathbb{E}_t[\theta^\top x])^2.
\end{aligned}$$

To proceed, we notice that $\tilde{Y}_{t,x}$ is a deterministic function of $Y_{t,x}$. Hence, $I(X_1^*; \tilde{Y}_{t,x}) \leq I(X_1^*; Y_{t,x})$. Then we have

$$g_t^\top \pi_t^{ts} \geq 2 \sum_{x, x' \in S_t^n} \mathbb{P}_t(X_{t,1}^* = x)\mathbb{P}_t(X_{t,1}^* = x')(\mathbb{E}_t[\theta^\top x \mid X_{t,1}^* = x'] - \mathbb{E}_t[\theta^\top x])^2$$

From here the same analysis for linear model can be applied. □

D Low-rank Matrix

We denote a policy at time t by $\pi_t \in [0,1]^{|\Omega_t^c|}$ each dimension corresponding to a unlabeled entry. Now we derive the instant regret and information gain. We denote an index by $a = (a_1, a_2)$, where a_1, a_2 are row and column indices.

Let μ be uniform distribution over Ω_t^c. Using Lemma 3 [34], we have

$$\begin{aligned}
\langle \pi_t, I_t \rangle &\geq \frac{2}{B^2+1} \sum_{a \in \Omega_t^c} \mu(a) \sum_{a^* \in \Omega_t^c} \mathbb{P}_t(X_{t,1}^* = a^*) \left(\mathbb{E}_t\left[Y_{t,a} \mid X_{t,1}^* = a^*\right] - \mathbb{E}_t\left[Y_{t,a}\right]\right)^2 \\
&= \frac{2}{B^2+1} \sum_{a^* \in \Omega_t^c} \mathbb{P}_t(X_{t,1}^* = a^*) \sum_{a \in \Omega_t^c} \mu(a) \left(\mathbb{E}_t\left[M_a \mid X_{t,1}^* = a^*\right] - \mathbb{E}_t\left[M_a\right]\right)^2 \\
&= \frac{2}{B^2+1} \sum_{a^* \in \Omega_t^c} \mathbb{P}_t(X_{t,1}^* = a^*) \sum_{a \in \Omega_t^c} \mu(a) \left(\mathbb{E}_t\left[M_a \mid X_{t,1}^* = a^*\right] - \mathbb{E}_t\left[M_a\right]\right)^2 \\
&= \frac{2}{B^2+1} \sum_{a^* \in \Omega_t^c} \mathbb{P}_t(X_{t,1}^* = a^*) \frac{1}{|\Omega_t^c|} \|M^{t,a^*}\|_F^2 \\
&\geq \frac{2}{B^2+1} \sum_{a^* \in \Omega_t^c} \mathbb{P}_t(X_{t,1}^* = a^*) \frac{1}{m^2} \|M^{t,a^*}\|_F^2 \\
&\geq \frac{2}{B^2+1} \sum_{a^* \in \Omega_t^c} \mathbb{P}_t(X_{t,1}^* = a^*) \frac{1}{Rm^2} \operatorname{trace}^2(M^{t,a^*}).
\end{aligned}$$

where $M^{t,a^*} := \mathbb{E}_t\left[M \mid X_{t,1}^* = a^*\right] - \mathbb{E}_t\left[M\right]$.

To upper bound the instant regret, let π_t^{ts} be the Thompson Sampling policy. We have

$$\begin{aligned}
\langle \pi_t, \Delta_t \rangle &= \sum_a \mathbb{P}_t(X_{t,1}^* = a) \left(\mathbb{E}_t \left[M_{X_1^*} \right] - \mathbb{E}_t \left[M_a \right] \right) \\
&= \sum_a \mathbb{P}_t(X_{t,1}^* = a) \left(\mathbb{E}_t \left[M_a \mid X_{t,1}^* = a \right] - \mathbb{E}_t \left[M_a \right] \right) \\
&\leq \sqrt{\sum_a \mathbb{P}_t(X_{t,1}^* = a) \left(\mathbb{E}_t \left[M_a \mid X_1^* = a \right] - \mathbb{E}_t \left[M_a \right] \right)^2} \\
&\leq \sqrt{\sum_a \mathbb{P}_t(X_{t,1}^* = a) \max_{a'} (M_{a'}^{t,a})^2} \\
&\quad \text{(Using Assumption 2)} \\
&\leq \sqrt{\sum_a \mathbb{P}_t(X_{t,1}^* = a) (4\frac{\gamma r}{m})^2 \operatorname{trace}^2(M^{t,a})}.
\end{aligned}$$

Now we consider a mixed policy: $\pi_t = p\mu + (1-p)\pi_t^{PS}$ for some $p \in [0,1]$.

$$\langle \pi_t, I_t \rangle \geq p \langle \mu, I_t \rangle \geq \frac{2p}{B^2+1} \sum_{a^* \in \Omega_t^c} \mathbb{P}_t(X_{t,1}^* = a^*) \frac{1}{Rm^2} \operatorname{trace}^2(M^{t,a}).$$

$$\begin{aligned}
\langle \pi_t, \Delta_t \rangle &\leq pB + (1-p) \left\langle \pi_t^{PS}, \Delta_t \right\rangle \\
&\leq pB + (1-p) \sqrt{\sum_a \mathbb{P}_t(X_{t,1}^* = a) (4\frac{\gamma r}{m})^2 \operatorname{trace}^2(M^{t,a})}
\end{aligned}$$

By optimizing p, we have

$$\frac{\langle \pi_t, \Delta_t \rangle^3}{\langle \pi_t, I_t \rangle} \leq 4B(B^2+1) r^3 \gamma^2.$$

Combined with Lemma 1, we have the following Bayesian regret bound

$$\mathcal{BR}(T, \text{IDS}) \lesssim (4B(B^2+1) r^3 \gamma^2 H(M) T^2)^{1/3}.$$

E Graph

Proof. The $\sqrt{T}$ term in the maximization can be achieved by replacing π_{IS} with π_{TS} and use the fact that each complete subgraph can be treated as a single node.

Let C_t be the smallest maximum independent set at the step t. To prove the $T^{2/3}$ term, we consider a mixture policy $\pi_t^{mix} = \gamma \pi_t^C + (1-\gamma)\pi_t^{TS}$ for some $\gamma > 0$, where π_t^C is the uniform distribution over C_t.

We have

$$\Delta_t(\pi_t^{mix}) \leq \gamma B + (1-\gamma)\Delta_t(\pi_t^{TS}).$$

For information gain, we have

$$g_t(\pi_t^{mix}) \geq \gamma g_t(\pi_C^t).$$

Since each node in S_t^n has an edge to at least one node in the maximum independent set (otherwise they have to be added to the set), we have

$$g_t(\pi_t^C) \geq \frac{1}{|C_t|} \sum_{x \in C_t} I_t(X_{t,1}^*; O_t(x)) \geq \frac{1}{|C_t|} \sum_{x \in S_t^n} I_t(X_{t,1}^*; Y_{t,x}).$$

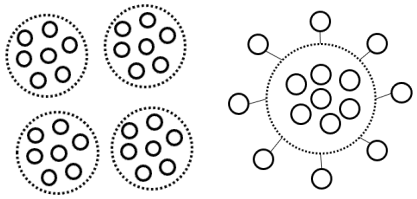


Figure 2: An illustration of graphs with rich structural information. Each dashed circle represent a complete subgraph. The graph on the left hand side has $\mathcal{X}(G) = 4$, while $N = 28$. The graph on the right hand side is a illustration of the star graph, in which each node outside the dashed circle has an edge to every node inside of the circle. This graph has $\mathcal{C}(G) = 1$, while $\mathcal{X}(G) = 8$ and $N = 15$.

Using (3) and Lemma 2 again, we have

$$\sum_{x \in S_t^n} I_t(X_{t,1}^*; Y_{t,x}) \geq \sum_{x \in S_t^n} \mathbb{P}_t(X_{t,1}^* = x) \left(\mathbb{E}_t \left[Y_{t,x} \mid X_{t,1}^* = x' \right] - \mathbb{E}_t \left[Y_{t,x} \right] \right)^2 .$$

We also have

$$(\Delta_t^\top \pi_t^{TS})^2 \leq \left(\sum_{x \in S_t^n} \mathbb{P}_t \left(X_{t,1}^* = x \right) \left(E_t \left[Y_{t,x} \mid X_{t,1}^* = x' \right] - \mathbb{E}_t \left[Y_{t,x} \right] \right) \right)^2 .$$

Therefore, we have

$$g_t(\pi_C^t) \geq \frac{1}{|C_t|} (\Delta_t(\pi_t^{TS}))^2 \geq \frac{1}{\mathcal{C}_T(G)} (\Delta_t(\pi_t^{TS}))^2.$$

Henceforth,

$$\begin{aligned} \Delta_t(\pi_t^{mix})^3 / g(\pi_t^{mix}) &\leq \frac{(\gamma B + (1-\gamma)(\mathcal{C}_T(G)^{1/2} g_t^{1/2}) / \gamma^{1/2})^3}{g_t} \\ &\leq \frac{(\gamma B + (\mathcal{C}_T(G)^{1/2} g_t^{1/2}) / \gamma^{1/2})^3}{g_t}. \end{aligned}$$

By optimizing γ, we have

$$\Delta_t(\pi_t^{mix})^3 / g(\pi_t^{mix}) \leq B\mathcal{C}(G).$$

□

F Generic results

The proof is analogous to the above proof for matrix and graph models. Consider a mixed policy $\tilde{\pi}_t = p\mu + (1-p)\pi_t^{ts}$.

We have

$$g_t^\top \tilde{\pi}_t \geq p g_t^\top \mu \geq p\phi(\Delta_t^\top \pi_t^{ts})^2$$

and

$$\Delta_t^\top \tilde{\pi}_t \leq (1-p)\Delta_t^T \pi_t^{ts} + pB$$

Thus

$$\frac{(\Delta_t^\top \tilde{\pi}_t)^3}{g_t^\top \tilde{\pi}_t} \leq \frac{((1-p)\Delta_t^T \pi_t^{ts} + pB)^3}{p\phi(\Delta_t^\top \pi_t^{ts})^2} \leq \frac{(\Delta_t^T \pi_t^{ts} + pB)^3}{p\phi(\Delta_t^\top \pi_t^{ts})^2}.$$

The proof is finished by optimizing p.

G Sparse linear model

Consider a linear regression problem

$$Y = X^\top \theta + \epsilon,$$

where $\theta \in \mathbb{R}^d$ and $\|\theta\|_0 = s$ is the sparsity and ϵ is the zero-mean noise. In general, we expect $d \gg T$, in which case, any dependence on d would lead to a linear regret.

Similarly to [14], we need to assume an exploratory unlabeled set.

Assumption 4. *Let $C_{min}(S^n) = \max_{\mu \in \mathcal{D}(S^n)} \sigma_{min}(\mathbb{E}_{x\sim\mu}[xx^\top])$. If $C_{min}(S_T^n) \geq 1$ almost surely for any algorithm, then we say that S^n is exploratory.*

Theorem 4 (Theorem 5.3 in [14])**.** *The following regret holds for IDS with $\lambda = 3$, if S^n is exploratory*

$$\mathcal{BR}(T, \text{IDS}) \lesssim \left(s^2 T^2 \Delta\right)^{1/3},$$

where

$$\Delta = \min\left(\log(n), 2s \log\left(CdT^{1/2}/s\right)\right)$$

for some constant $C > 0$.

The proof in [14] on sparse linear bandit is also applicable here. The only difference is that we make a stronger assumption on the exploratory set stating that the unlabeled dataset is still exploratory after eliminating any T elements. This is to guarantee the exploratory set for any step during the decision-making process.

H Experiments

H.1 Approximate algorithm

The approximate algorithm, SampleVIDS is given in Algorithm 2.

Algorithm 2 SampleVIDS (Sample Variance-based IDS)

Input: Unlabeled dataset S_t^n, prior distribution ϕ, total number of steps T, number of posterior samples M, constant λ.
Initialize history $\mathcal{F}_0 = \{\}$.
for $t = 1$ **to** T **do**
  Sample $\theta_1, \ldots, \theta_M$ from $\phi(\cdot \mid \mathcal{F}_{t-1})$.
  Calculate instant regret by $\Delta_t(x) = \sum_{i=1}^M \max_{x' \in S_t^n} f_{\theta_i}(x') - f_{\theta_i}(x)$ for all $x \in S_t^n$.
  Let $\Theta(x) = \{\theta_i : x \in \arg\max_{x' \in S_t^n} f_{\theta_i}(x')\}$ and $\bar{f}(x) = \sum_{i=1}^M f_{\theta_i}(x)/M$.
  Calculate variance-based information ratio for all $x \in S_t^n$ by

$$v_t(x) = \sum_{x' \in S_t^n} \frac{|\Theta(x')|}{M} \left(\frac{1}{|\Theta(x')|} \sum_{\theta \in \Theta(x')} f_\theta(x) - \bar{f}(x)\right)^2.$$

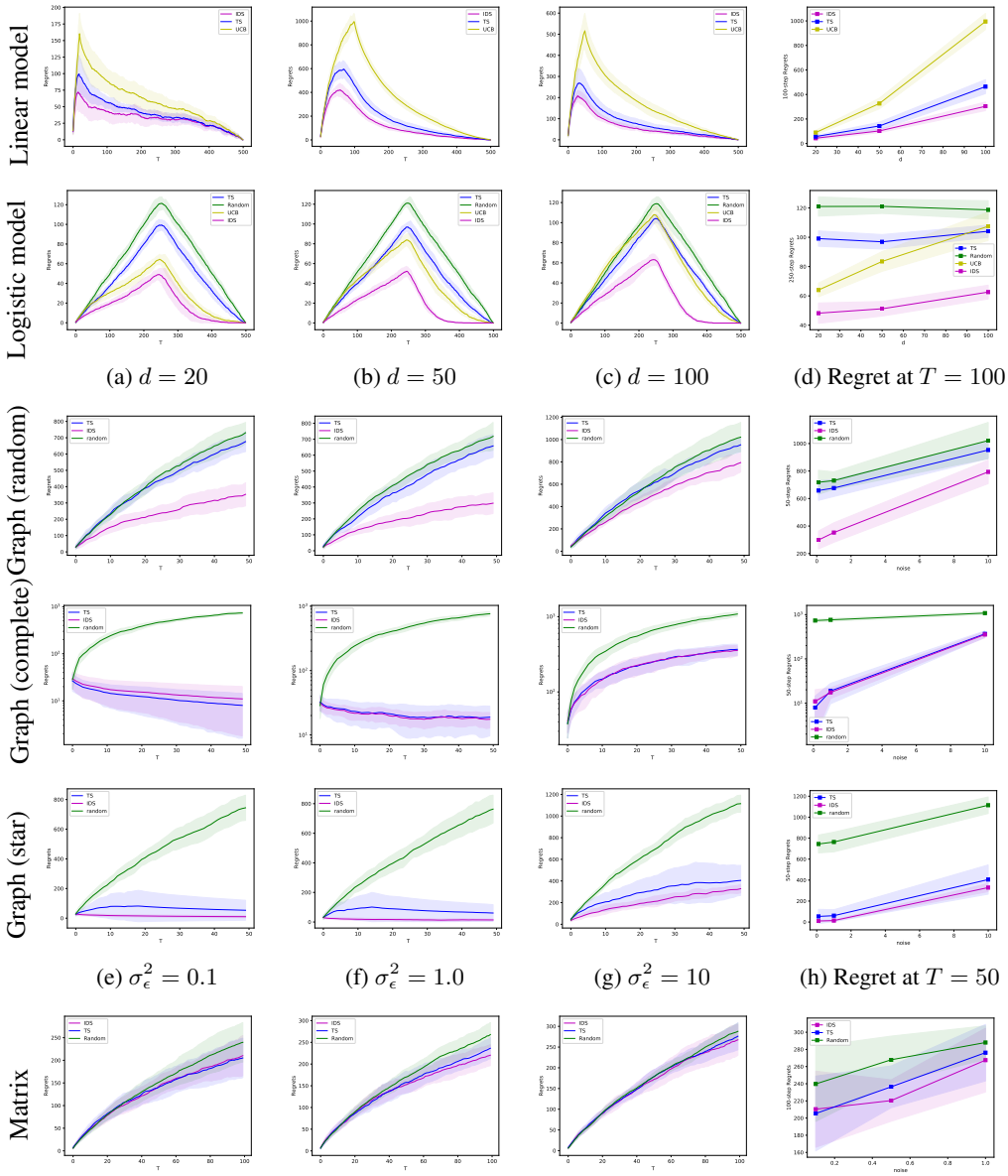
  Calculate the information ratio $\Psi_t(x) = \Delta_t(x)^\lambda / v_t(x)$ and label $X_t = \arg\max_x \Psi_t(x)$.
  Update history $\mathcal{F}_t = \mathcal{F}_{t-1} \cup \{(X_t, Y_t)\}$.
end for

H.2 Complete graph for simulation studies

We provide the complete simulation results in Figure 3.

H.3 Additional information for reaction condition discovery

Additional information on two datasets. The complete library of informer molecules for Photoredox Nickel Dual-Catalysis (PNDC) and the structure of reactions for C-N Cross-Coupling with Isoxazoles (CNCCI) are provided in Figure 4 and 5. For PNDC, note that we only report the results



(a) $d = 20$ (b) $d = 50$ (c) $d = 100$ (d) Regret at $T = 100$

(e) $\sigma_\epsilon^2 = 0.1$ (f) $\sigma_\epsilon^2 = 1.0$ (g) $\sigma_\epsilon^2 = 10$ (h) Regret at $T = 50$

(i) $\sigma_\epsilon^2 = 0.1$ (j) $\sigma_\epsilon^2 = 0.5$ (k) $\sigma_\epsilon^2 = 1$ (l) Regret at $T = 100$

Figure 3: Cumulative regret curves for linear regression (row 1), logistic regression (row 2), random graph, star graph (row 3, 4) and low-rank matrix (row 5). The left three columns are the regret curves for different dimensions $d = 20, 50, 100$ in linear and logistic regression simulations and for different noise levels in graph and low-rank matrix simulations. The last column is the early-stage cumulative regret. The confidence ranges are given by the standard deviation of 10 independent runs.

for X2, X3, X4, X5, X6, X8, X11, X12, X13, X14, X15 in PNDC due to very low yields across all reaction conditions for the remaining molecules. Additionally, while the original dataset considers 12 photocatalysts (total 96 reaction conditions) for each molecule, we consider 10 photocatalysts (80 reaction conditions), due to the unavailability of descriptors that we intended to use. For CNCCI, only pairs of catalyst and base with datapoints of all 330 combinations of 15 aryl halides and 22 isoxazoles were considered. The features used for CNCCI is a subset of those prepared in [2], with highest feature importance values in models presented in the original work.

The two datasets are included in **PNDC.xlsx** and **CNCCI.xlsx** under data folder in the code. Each file has two sheets for yield and descriptors respectively.

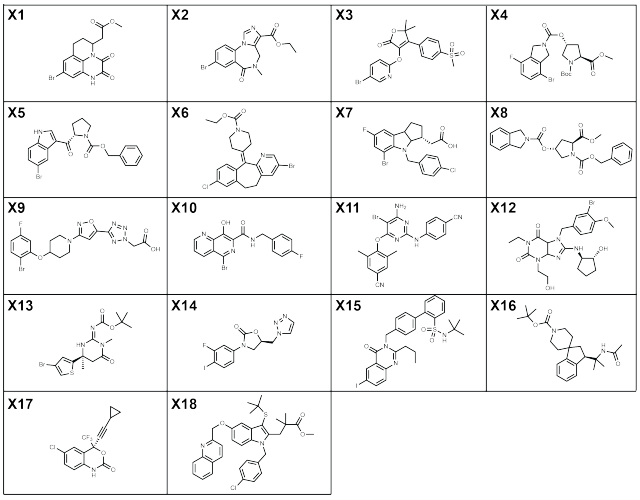


Figure 4: Complete library for PNDC (Figure 2 in [11]).

Response distribution. Figure 6 provides the distributions of the response variables in two dataset.

Complete regret curves for PNDC. The complete regret curves for PNDC is given in Figure 7.

Complete regret curves for CNCCI. The complete regret curves for CNCCI is given in Figure 8.

I Comparison to ENS

In this section, we briefly compare IDS with ENS (efficient nonmyopic search).

We observed that ENS tends to over explore in the experiments on linear models. See our new Figure 9 in the appendix. We believe this is because ENS assumes that the labels of all remaining unlabeled points are conditionally independent. That is the extra gain by observing the new label y_t is uniform across all the remaining $T - t$ points. This is over estimating the gain, because in the later stage when the estimates on $\Pr(y = 1 \mid x, \mathcal{F}_t)$ are more accurate the extra gain from observing a single label is also much less. ENS thus weighs too much on the exploration side. We highly believe that this will lead to linear regret instead of T regret that can be achieved by IDS. This is also reflected in Figure 9, where ENS performs worse than IDS when noise level of the problem is low and we need more exploitation. In general, IDS provides a more flexible balance between exploration and exploitation.

We compare IDS with ENS on linear models with different level of noise. In general, a more noisy model requires more exploration. In Figure 9, ENS performs worse in the low-noise models while outperform IDS in higher noise settings.

I.1 Computation resources and implementation assets.

All the computation are done on MacBook Pro with 1.4 GHz Quad-Core Intel Core i5 Processor and 16GB Memory. Part of the code is from [39].

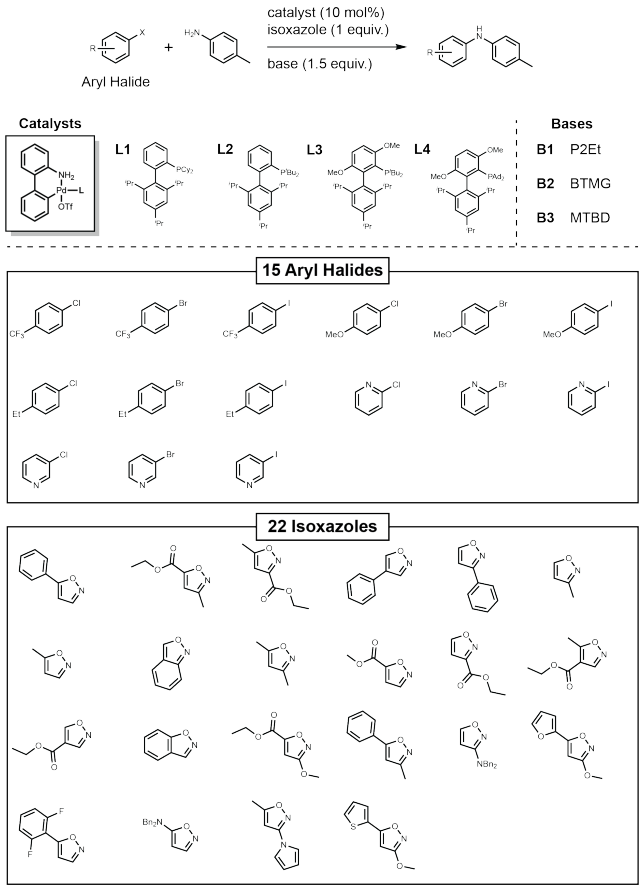
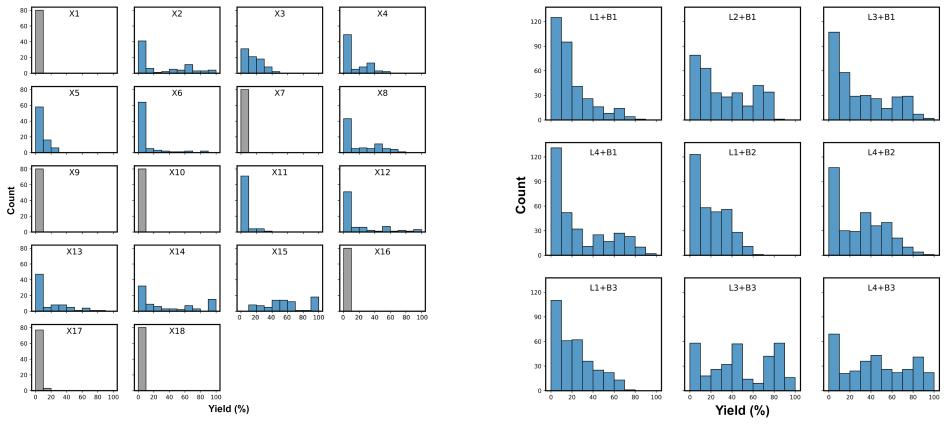


Figure 5: Complete library for CNCCI (Figure 3 in [2]).



(a) PNDC (b) CNCCI

Figure 6: The distribution of the response variables in PNDC and CNCCI dataset. x-axes are response variables (yield rate) and each panel corresponds to one target molecule.

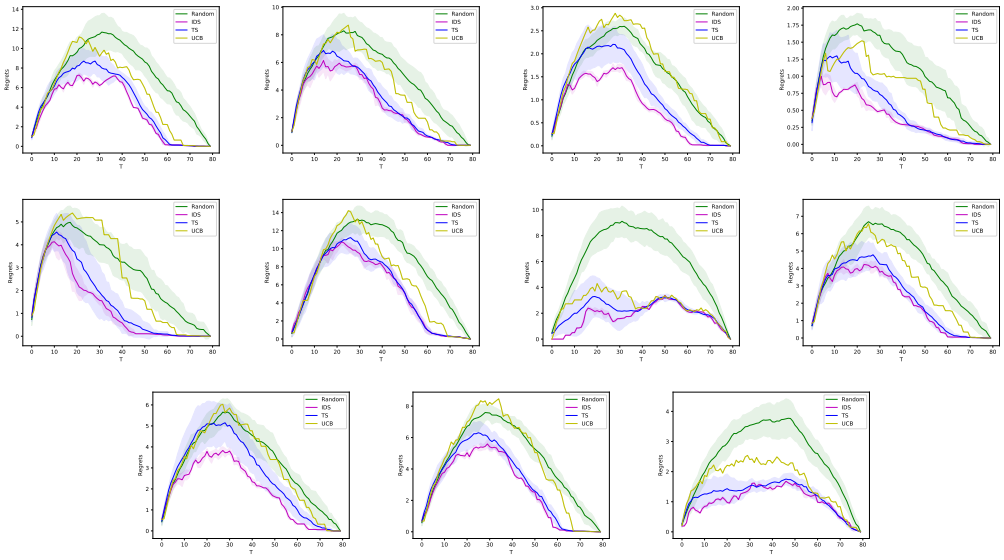


Figure 7: The whole horizon regret curves for PNDC dataset that corresponds to Table 1. The confidence interval are the standard deviation calculated from 10 independent runs.

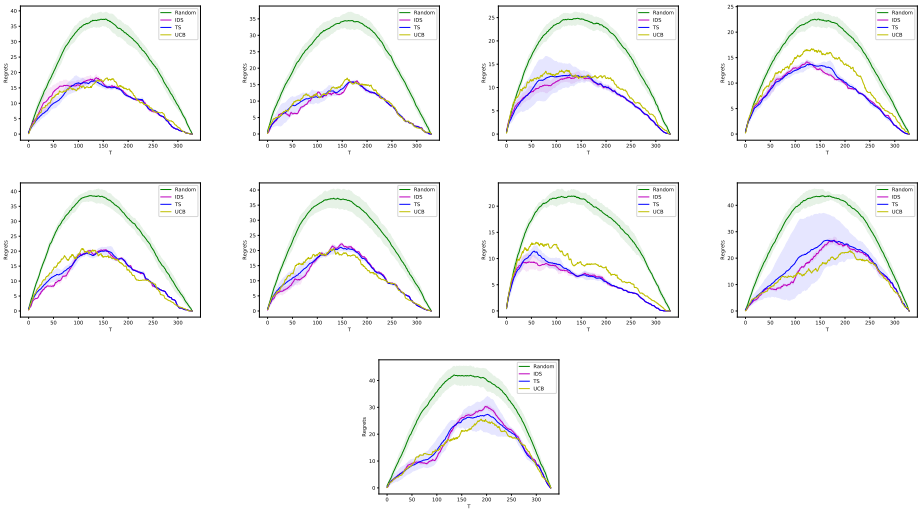


Figure 8: The whole horizon regret curves for CNCCI dataset that corresponds to Table 1. The confidence interval are the standard deviation calculated from 10 independent runs.

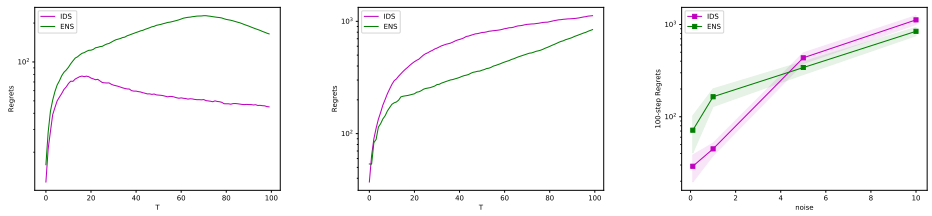


Figure 9: Comparing IDS with ENS on linear models with $\sigma = 0.1, 1, 5, 10$.