Fast Rates for Bandit PAC Multiclass Classification

Liad Erez

Joint work with: Alon Cohen, Tomer Koren, Yishay Mansour, Shay Moran

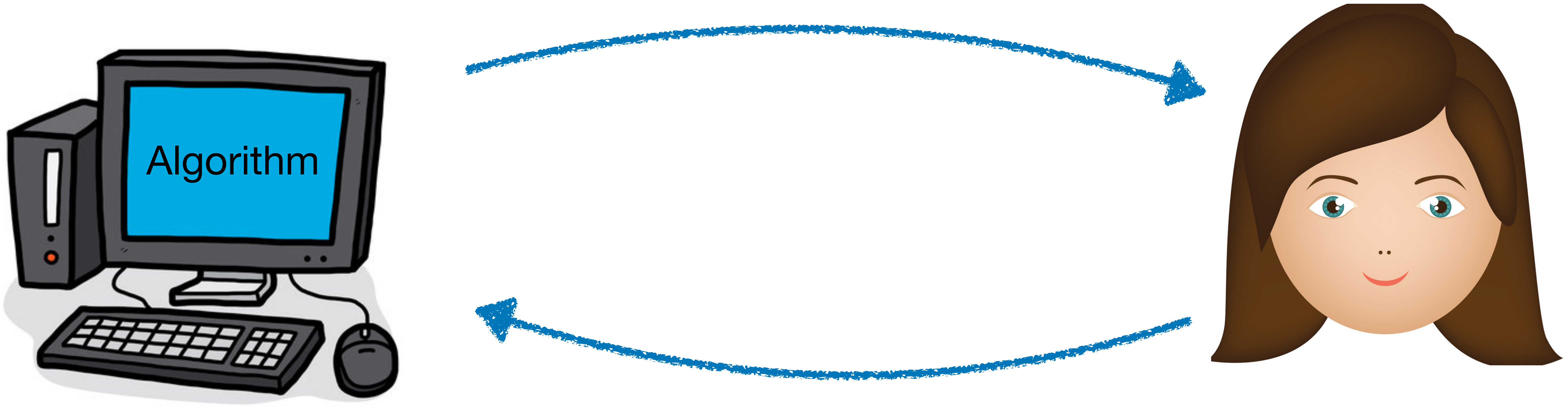
European Research Council
Established by the European Commission

Blavatnik School of Computer Science
Raymond & Beverly Sackler
Faculty of Exact Sciences
Tel Aviv University

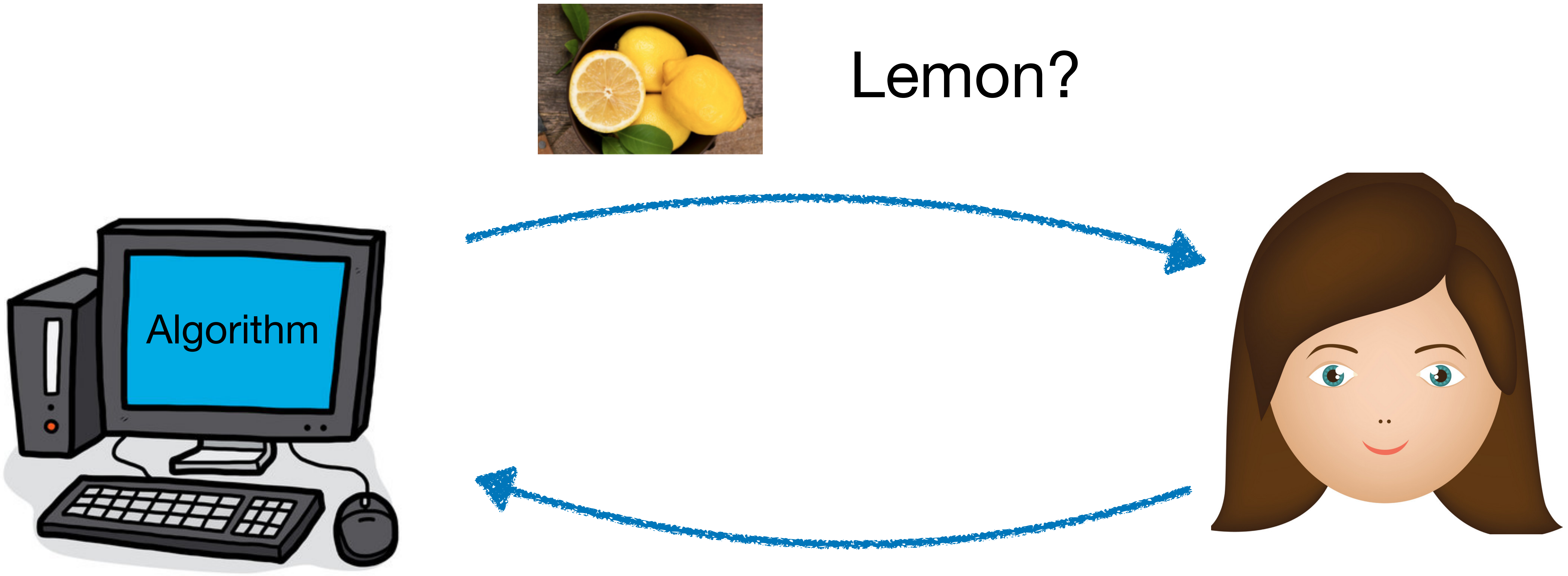
Google

Bandit Multiclass Classification

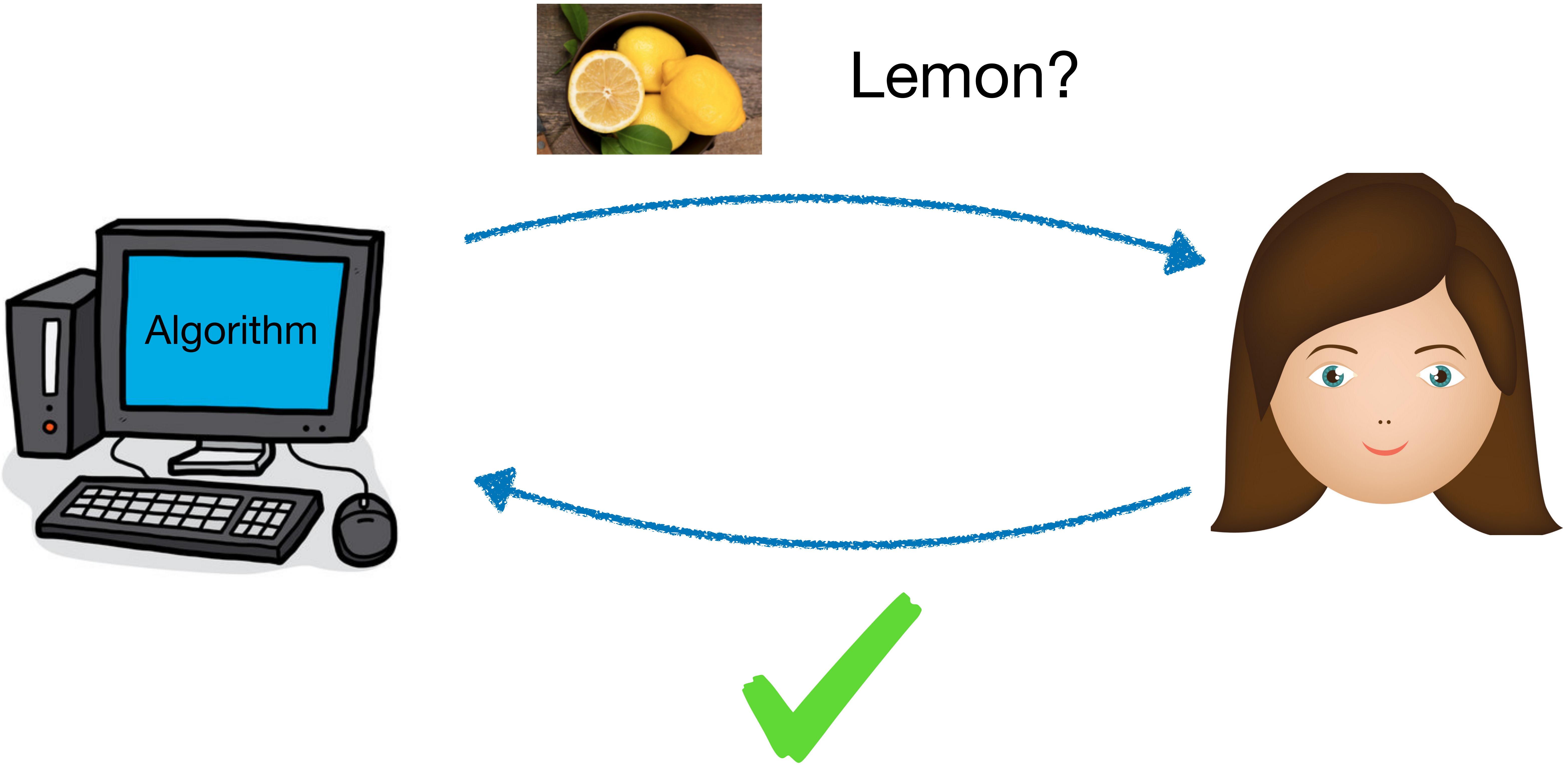
Bandit Multiclass Classification



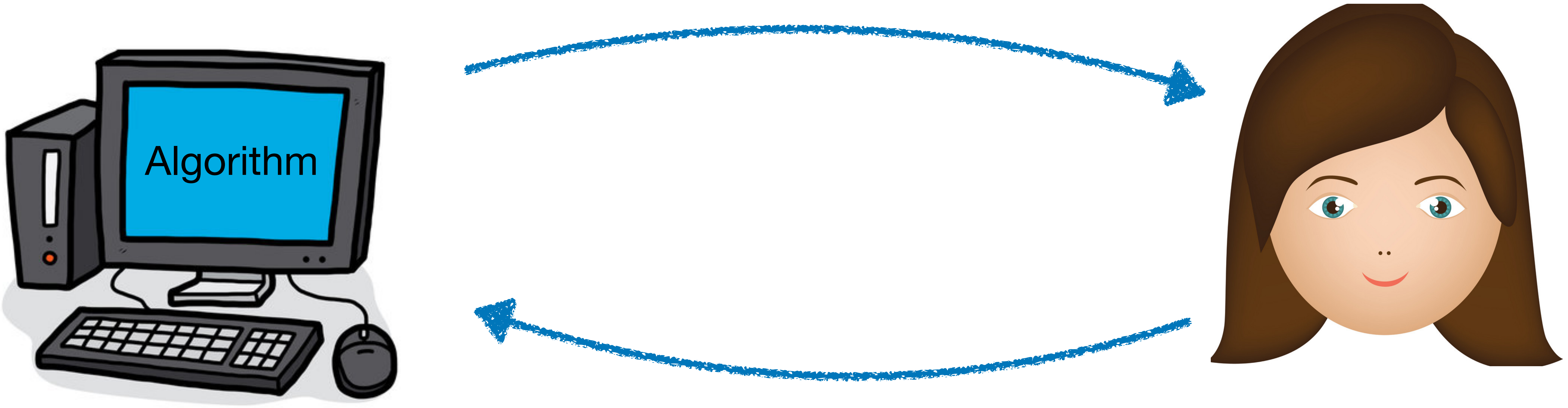
Bandit Multiclass Classification



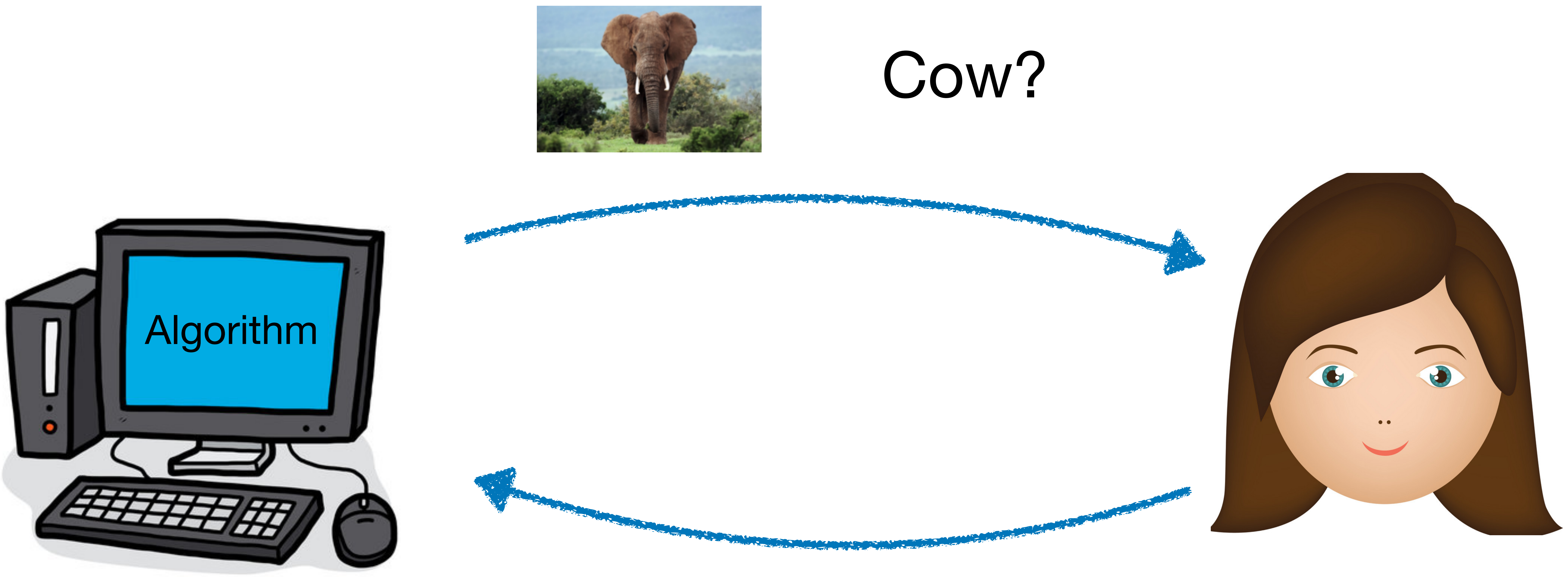
Bandit Multiclass Classification



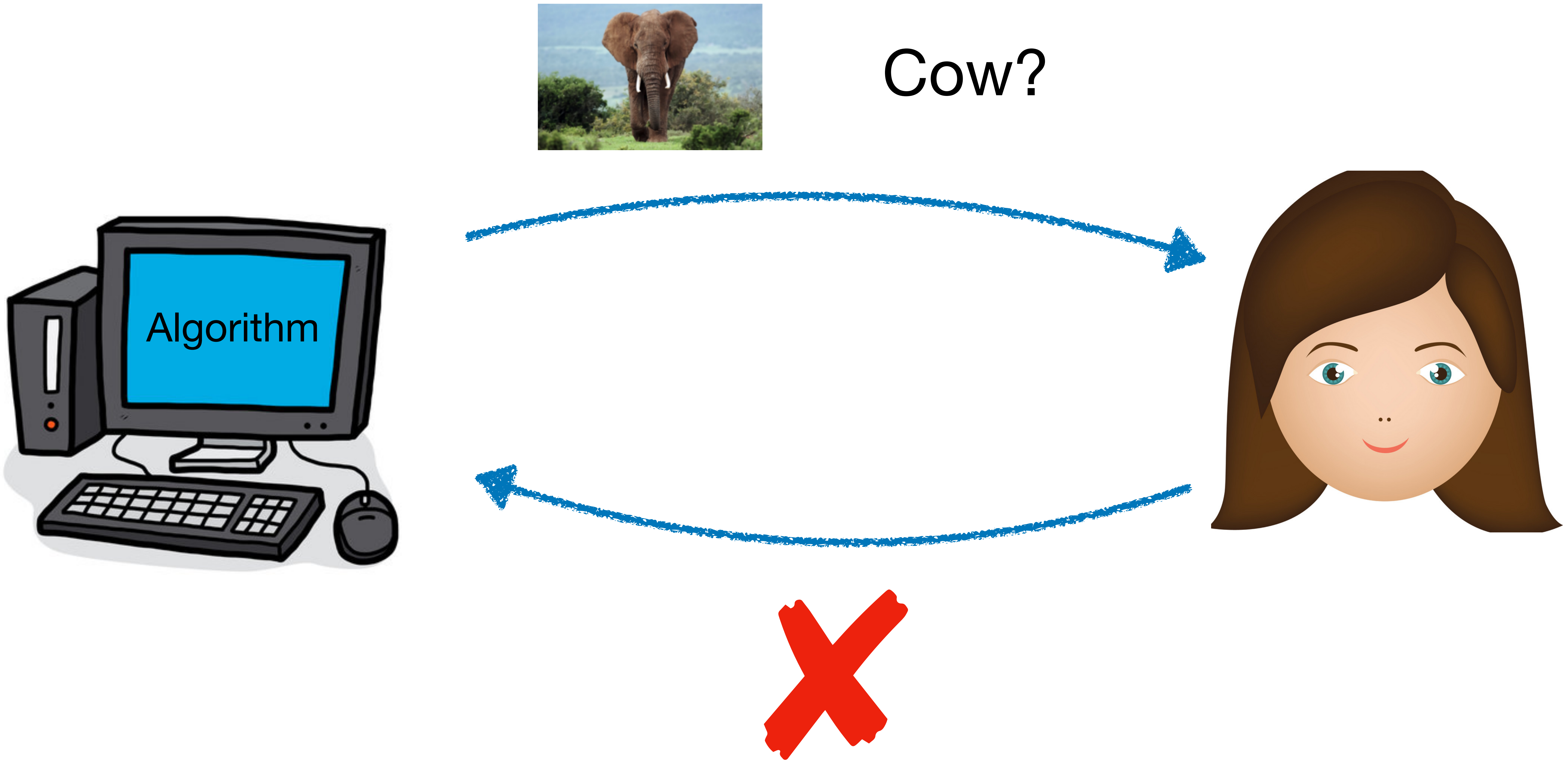
Bandit Multiclass Classification



Bandit Multiclass Classification



Bandit Multiclass Classification



Problem Setup

(Agnostic) PAC bandit multiclass classification

Problem Setup

(Agnostic) PAC bandit multiclass classification

Domain $\mathcal{X}$, label space $\mathcal{Y}$ with $|\mathcal{Y}| = K$, hypothesis class $\mathcal{H} \subseteq \{\mathcal{X} \to \mathcal{Y}\}$, (unknown) distribution $\mathcal{D}$ over $\mathcal{X} \times \mathcal{Y}$.

Problem Setup

(Agnostic) PAC bandit multiclass classification

Domain $\mathcal{X}$, label space $\mathcal{Y}$ with $|\mathcal{Y}| = K$, hypothesis class $\mathcal{H} \subseteq \{\mathcal{X} \to \mathcal{Y}\}$, (unknown) distribution $\mathcal{D}$ over $\mathcal{X} \times \mathcal{Y}$.

For $i = 1,2,3,\ldots$:

Problem Setup

(Agnostic) PAC bandit multiclass classification

Domain $\mathcal{X}$, label space $\mathcal{Y}$ with $|\mathcal{Y}| = K$, hypothesis class $\mathcal{H} \subseteq \{\mathcal{X} \to \mathcal{Y}\}$, (unknown) distribution $\mathcal{D}$ over $\mathcal{X} \times \mathcal{Y}$.

For $i = 1,2,3,\dots$:

- Environment generates $(x_i, y_i) \sim \mathcal{D}$, x_i is revealed to the learner;

Problem Setup

(Agnostic) PAC bandit multiclass classification

Domain $\mathcal{X}$, label space $\mathcal{Y}$ with $|\mathcal{Y}| = K$, hypothesis class $\mathcal{H} \subseteq \{\mathcal{X} \to \mathcal{Y}\}$, (unknown) distribution $\mathcal{D}$ over $\mathcal{X} \times \mathcal{Y}$.

For $i = 1,2,3,\ldots$:

- Environment generates $(x_i, y_i) \sim \mathcal{D}$, x_i is revealed to the learner;
- Learner predicts $\hat{y}_i \in \mathcal{Y}$;

Problem Setup

(Agnostic) PAC bandit multiclass classification

Domain $\mathcal{X}$, label space $\mathcal{Y}$ with $|\mathcal{Y}| = K$, hypothesis class $\mathcal{H} \subseteq \{\mathcal{X} \to \mathcal{Y}\}$, (unknown) distribution $\mathcal{D}$ over $\mathcal{X} \times \mathcal{Y}$.

For $i = 1,2,3,\ldots$:

- Environment generates $(x_i, y_i) \sim \mathcal{D}$, x_i is revealed to the learner;
- Learner predicts $\hat{y}_i \in \mathcal{Y}$;
- Learner observes whether or not prediction $\hat{y}_i$ is correct, namely $\mathbf{1}\{\hat{y}_i = y_i\}$ (bandit feedback).

Problem Setup

(Agnostic) PAC bandit multiclass classification

Domain $\mathscr{X}$, label space $\mathscr{Y}$ with $|\mathscr{Y}| = K$, hypothesis class $\mathscr{H} \subseteq \{\mathscr{X} \to \mathscr{Y}\}$, (unknown) distribution $\mathscr{D}$ over $\mathscr{X} \times \mathscr{Y}$.

For $i = 1,2,3,\ldots$:

- Environment generates $(x_i, y_i) \sim \mathscr{D}$, x_i is revealed to the learner;
- Learner predicts $\hat{y}_i \in \mathscr{Y}$;
- Learner observes whether or not prediction $\hat{y}_i$ is correct, namely $\mathbf{1}\{\hat{y}_i = y_i\}$ (bandit feedback).

Objective (PAC learning): Given $\epsilon, \delta > 0$, learn $\hat{h} \in \mathscr{H}$ such that w.p. at least $1 - \delta$:

Problem Setup

(Agnostic) PAC bandit multiclass classification

Domain $\mathscr{X}$, label space $\mathscr{Y}$ with $|\mathscr{Y}| = K$, hypothesis class $\mathscr{H} \subseteq \{\mathscr{X} \to \mathscr{Y}\}$, (unknown) distribution $\mathscr{D}$ over $\mathscr{X} \times \mathscr{Y}$.

For $i = 1,2,3,\ldots$:

- Environment generates $(x_i, y_i) \sim \mathscr{D}$, x_i is revealed to the learner;
- Learner predicts $\hat{y}_i \in \mathscr{Y}$;
- Learner observes whether or not prediction $\hat{y}_i$ is correct, namely $\mathbf{1}\{\hat{y}_i = y_i\}$ (bandit feedback).

<u>**Objective (PAC learning):**</u> Given $\epsilon, \delta > 0$, learn $\hat{h} \in \mathscr{H}$ such that w.p. at least $1 - \delta$:

$$L_{\mathscr{D}}(\hat{h}) - L_{\mathscr{D}}(h) \leq \epsilon \quad \forall h \in \mathscr{H},$$

Problem Setup

(Agnostic) PAC bandit multiclass classification

Domain $\mathcal{X}$, label space $\mathcal{Y}$ with $|\mathcal{Y}| = K$, hypothesis class $\mathcal{H} \subseteq \{\mathcal{X} \to \mathcal{Y}\}$, (unknown) distribution $\mathcal{D}$ over $\mathcal{X} \times \mathcal{Y}$.

For $i = 1,2,3,\ldots$:

- Environment generates $(x_i, y_i) \sim \mathcal{D}$, x_i is revealed to the learner;
- Learner predicts $\hat{y}_i \in \mathcal{Y}$;
- Learner observes whether or not prediction $\hat{y}_i$ is correct, namely $\mathbf{1}\{\hat{y}_i = y_i\}$ (bandit feedback).

Objective (PAC learning): Given $\epsilon, \delta > 0$, learn $\hat{h} \in \mathcal{H}$ such that w.p. at least $1 - \delta$:

$$L_{\mathcal{D}}(\hat{h}) - L_{\mathcal{D}}(h) \leq \epsilon \quad \forall h \in \mathcal{H},$$

where $L_{\mathcal{D}}(h) := \Pr_{(x,y)\sim\mathcal{D}}[h(x) \neq y]$ is the zero-one population loss of h.

Problem Setup

(Agnostic) PAC bandit multiclass classification

Domain $\mathcal{X}$, label space $\mathcal{Y}$ with $|\mathcal{Y}| = K$, hypothesis class $\mathcal{H} \subseteq \{\mathcal{X} \to \mathcal{Y}\}$, (unknown) distribution $\mathcal{D}$ over $\mathcal{X} \times \mathcal{Y}$.

For $i = 1,2,3,\ldots$:

- Environment generates $(x_i, y_i) \sim \mathcal{D}$, x_i is revealed to the learner;
- Learner predicts $\hat{y}_i \in \mathcal{Y}$;
- Learner observes whether or not prediction $\hat{y}_i$ is correct, namely $\mathbf{1}\{\hat{y}_i = y_i\}$ (bandit feedback).

Objective (PAC learning): Given $\epsilon, \delta > 0$, learn $\hat{h} \in \mathcal{H}$ such that w.p. at least $1 - \delta$:

$$L_{\mathcal{D}}(\hat{h}) - L_{\mathcal{D}}(h) \leq \epsilon \quad \forall h \in \mathcal{H},$$

where $L_{\mathcal{D}}(h) := \Pr_{(x,y)\sim\mathcal{D}}[h(x) \neq y]$ is the zero-one population loss of h.

Performance is measured by *sample complexity*: # of samples required for PAC guarantee.

Known Results

Known Results

- A naive approach allows for a sample complexity of $\widetilde{O}\left(K/\epsilon^2\right)$* by sampling labels at random and estimating the expected rewards of all hypotheses in $\mathscr{H}$.

* Here and henceforth we omit $\log(|\mathscr{H}|/\delta)$ factors

Known Results

- A naive approach allows for a sample complexity of $\widetilde{O}\left(K/\epsilon^2\right)$* by sampling labels at random and estimating the expected rewards of all hypotheses in $\mathscr{H}$.

* Here and henceforth we omit $\log(|\mathscr{H}|/\delta)$ factors

Known Results

- A naive approach allows for a sample complexity of $\widetilde{O}\left(K/\epsilon^2\right)$* by sampling labels at random and estimating the expected rewards of all hypotheses in $\mathscr{H}$.

- Such rates can be obtained while also minimizing *regret* using efficient algorithms (e.g. Dudik et al. '11, Agrawal et al. '14) with regret bounds of $\widetilde{O}\left(\sqrt{KT}\right)$ for *contextual bandits*.

* Here and henceforth we omit $\log(|\mathscr{H}|/\delta)$ factors

Known Results

- A naive approach allows for a sample complexity of $\widetilde{O}\left(K/\epsilon^2\right)$* by sampling labels at random and estimating the expected rewards of all hypotheses in $\mathscr{H}$.

- Such rates can be obtained while also minimizing *regret* using efficient algorithms (e.g. Dudik et al. '11, Agrawal et al. '14) with regret bounds of $\widetilde{O}\left(\sqrt{KT}\right)$ for *contextual bandits*.

- Bandit multiclass classification is a case of contextual bandits with sparse rewards. Despite sparsity, the optimal regret for bandit multi class classification is lower bounded by $\Omega(\sqrt{KT})$ (Erez et al. '24).

* Here and henceforth we omit $\log(|\mathscr{H}|/\delta)$ factors

Known Results

- A naive approach allows for a sample complexity of $\widetilde{O}\left(K/\epsilon^2\right)$* by sampling labels at random and estimating the expected rewards of all hypotheses in $\mathscr{H}$.

- Such rates can be obtained while also minimizing *regret* using efficient algorithms (e.g. Dudik et al. '11, Agrawal et al. '14) with regret bounds of $\widetilde{O}\left(\sqrt{KT}\right)$ for *contextual bandits*.

- Bandit multiclass classification is a case of contextual bandits with sparse rewards. Despite sparsity, the optimal regret for bandit multi class classification is lower bounded by $\Omega(\sqrt{KT})$ (Erez et al. '24).

Question: Is it possible to guarantee rates faster than K/ϵ^2 for bandit PAC multiclass classification using an efficient algorithm?

Our Main Result

Our Main Result

Theorem: There is an efficient* algorithm which satisfies the (ϵ, δ)-PAC guarantee for bandit multiclass classification using a sample complexity of

Our Main Result

<u>Theorem:</u> There is an efficient* algorithm which satisfies the (ϵ, δ)-PAC guarantee for bandit multiclass classification using a sample complexity of

* Given access to a weighted ERM oracle for $\mathscr{H}$

Our Main Result

Theorem: There is an efficient* algorithm which satisfies the (ϵ, δ)-PAC guarantee for bandit multiclass classification using a sample complexity of

$$\widetilde{O}\left(\left(K^9 + \frac{1}{\epsilon^2}\right)\log\frac{|\mathcal{H}|}{\delta}\right)$$

* Given access to a weighted ERM oracle for $\mathcal{H}$

Our Main Result

Theorem: There is an efficient* algorithm which satisfies the (ϵ, δ)-PAC guarantee for bandit multiclass classification using a sample complexity of

$\log |\mathscr{H}|$ can be replaced by d for classes of finite Natarajan dimension d

$$\widetilde{O}\left(\left(K^9 + \frac{1}{\epsilon^2}\right) \log \frac{|\mathscr{H}|}{\delta}\right)$$

* Given access to a weighted ERM oracle for $\mathscr{H}$

Our Main Result

Theorem: There is an efficient* algorithm which satisfies the (ϵ, δ)-PAC guarantee for bandit multiclass classification using a sample complexity of

$\log|\mathcal{H}|$ can be replaced by d for classes of finite Natarajan dimension d

$$\widetilde{O}\left(\left(K^9 + \frac{1}{\epsilon^2}\right)\log\frac{|\mathcal{H}|}{\delta}\right)$$

- In the regime where $\epsilon \ll K^{-4}$, this rate is asymptotically faster than K/ϵ^2.

* Given access to a weighted ERM oracle for $\mathcal{H}$

Our Main Result

Theorem: There is an efficient* algorithm which satisfies the (ϵ, δ)-PAC guarantee for bandit multiclass classification using a sample complexity of

$\log|\mathscr{H}|$ can be replaced by d for classes of finite Natarajan dimension d

$$\widetilde{O}\left(\left(K^9 + \frac{1}{\epsilon^2}\right)\log\frac{|\mathscr{H}|}{\delta}\right)$$

- In the regime where $\epsilon \ll K^{-4}$, this rate is asymptotically faster than K/ϵ^2.
- Nearly matches the *full-information* rate of $1/\epsilon^2$!

* Given access to a weighted ERM oracle for $\mathscr{H}$

Our Approach (finite case)

Our Approach (finite case)

Our algorithm operates in two phases:

Our Approach (finite case)

Our algorithm operates in two phases:

Phase 1: Predict using $\approx K^9$ random labels and compute an exploration distribution $\hat{P} \in \Delta_{\mathcal{H}}$ satisfying:

Our Approach (finite case)

Our algorithm operates in two phases:

Phase 1: Predict using $\approx K^9$ random labels and compute an exploration distribution $\hat{P} \in \Delta_{\mathscr{H}}$ satisfying:

$$\mathbb{E}_{(x,y)\sim\mathscr{D}}\left[\frac{\mathbb{I}\{h(x)=y\}}{W^{\gamma}_{x,y}(\hat{P})}\right] \leq C \quad \forall h \in \mathscr{H},$$

Our Approach (finite case)

Our algorithm operates in two phases:

<u>Phase 1:</u> Predict using $\approx K^9$ random labels and compute an exploration distribution $\hat{P} \in \Delta_{\mathscr{H}}$ satisfying:

$$\mathbb{E}_{(x,y)\sim\mathscr{D}}\left[\frac{\mathbb{I}\{h(x)=y\}}{W^{\gamma}_{x,y}(\hat{P})}\right] \leq C \quad \forall h \in \mathscr{H},$$

where $W^{\gamma}_{x,y}(\hat{P}) := (1-\gamma)\sum_{h\in\mathscr{H}} \hat{P}(h)\mathbb{I}\{h(x)=y\} + \gamma/K$, and C is an absolute constant and $\gamma > 0$ is a parameter.

Our Approach (finite case)

Our algorithm operates in two phases:

Phase 1: Predict using $\approx K^9$ random labels and compute an exploration distribution $\hat{P} \in \Delta_{\mathscr{H}}$ satisfying:

$$\mathbb{E}_{(x,y)\sim\mathscr{D}}\left[\frac{\mathbb{I}\{h(x)=y\}}{W^{\gamma}_{x,y}(\hat{P})}\right] \le C \quad \forall h \in \mathscr{H},$$

where $W^{\gamma}_{x,y}(\hat{P}) := (1-\gamma)\sum_{h\in\mathscr{H}} \hat{P}(h)\mathbb{I}\{h(x)=y\} + \gamma/K$, and C is an absolute constant and $\gamma > 0$ is a parameter.

(**Intuition:** Importance weighted reward estimator induced by $\hat{P}$ has variance bounded by C.)

Our Approach (finite case)

Our algorithm operates in two phases:

Phase 1: Predict using $\approx K^9$ random labels and compute an exploration distribution $\hat{P} \in \Delta_{\mathcal{H}}$ satisfying:

$$\mathbb{E}_{(x,y)\sim\mathcal{D}}\left[\frac{\mathbb{I}\{h(x)=y\}}{W^{\gamma}_{x,y}(\hat{P})}\right] \leq C \quad \forall h \in \mathcal{H},$$

Achieved via stochastic Frank-Wolfe optimization procedure

where $W^{\gamma}_{x,y}(\hat{P}) := (1-\gamma)\sum_{h\in\mathcal{H}} \hat{P}(h)\mathbb{I}\{h(x)=y\} + \gamma/K$, and C is an absolute constant and $\gamma > 0$ is a parameter.

(**Intuition:** Importance weighted reward estimator induced by $\hat{P}$ has variance bounded by C.)

Our Approach (finite case)

Our algorithm operates in two phases:

Phase 1: Predict using $\approx K^9$ random labels and compute an exploration distribution $\hat{P} \in \Delta_{\mathscr{H}}$ satisfying:

$$\mathbb{E}_{(x,y)\sim\mathscr{D}}\left[\frac{\mathbb{I}\{h(x)=y\}}{W^{\gamma}_{x,y}(\hat{P})}\right] \leq C \quad \forall h \in \mathscr{H},$$

Achieved via stochastic Frank-Wolfe optimization procedure

where $W^{\gamma}_{x,y}(\hat{P}) := (1-\gamma)\sum_{h\in\mathscr{H}} \hat{P}(h)\mathbb{I}\{h(x)=y\} + \gamma/K$, and C is an absolute constant and $\gamma > 0$ is a parameter.

(**Intuition:** Importance weighted reward estimator induced by $\hat{P}$ has variance bounded by C.)

Phase 2: Sample $\approx 1/\epsilon^2$ hypotheses from $\hat{P}$ and output the hypothesis with highest estimated reward.

Our Approach (finite case)

Our algorithm operates in two phases:

Phase 1: Predict using $\approx K^9$ random labels and compute an exploration distribution $\hat{P} \in \Delta_{\mathscr{H}}$ satisfying:

$$\mathbb{E}_{(x,y)\sim\mathscr{D}}\left[\frac{\mathbb{I}\{h(x)=y\}}{W^{\gamma}_{x,y}(\hat{P})}\right] \leq C \quad \forall h \in \mathscr{H},$$

Achieved via stochastic Frank-Wolfe optimization procedure

where $W^{\gamma}_{x,y}(\hat{P}) := (1-\gamma)\sum_{h\in\mathscr{H}} \hat{P}(h)\mathbb{I}\{h(x)=y\} + \gamma/K$, and C is an absolute constant and $\gamma > 0$ is a parameter.

(**Intuition:** Importance weighted reward estimator induced by $\hat{P}$ has variance bounded by C.)

Phase 2: Sample $\approx 1/\epsilon^2$ hypotheses from $\hat{P}$ and output the hypothesis with highest estimated reward.

(**Intuition:** Bernstein's inequality guarantees that $\approx 1/\epsilon^2$ samples suffice in order to uniformly estimates the rewards for all hypotheses in $\mathscr{H}$.)

Thank You!