On the Optimality of the Median-of-Means Estimator under Adversarial Contamination

Xabier de Juan Santiago Mazuelas
Basque Center for Applied Mathematics

Conference on Neural Information Processing Systems
NeurIPS 2025

EUSKO JAURLARITZA
GOBIERNO VASCO

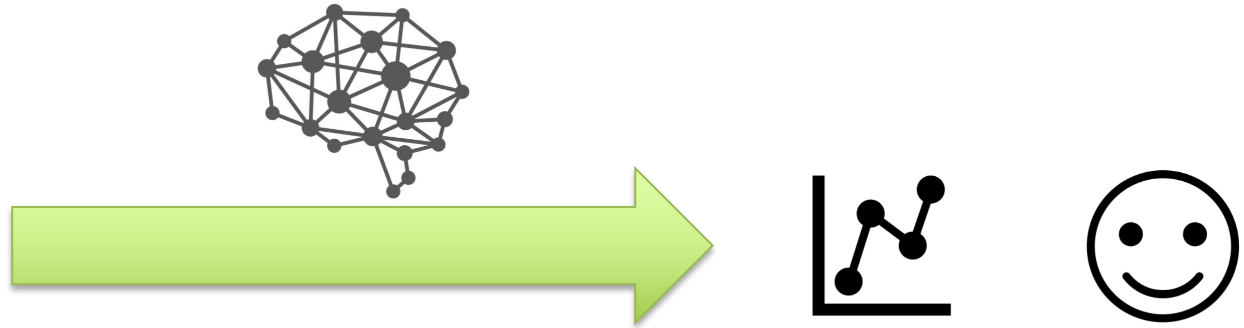
basque center for applied mathematics

EXCELENCIA
SEVERO
OCHOA

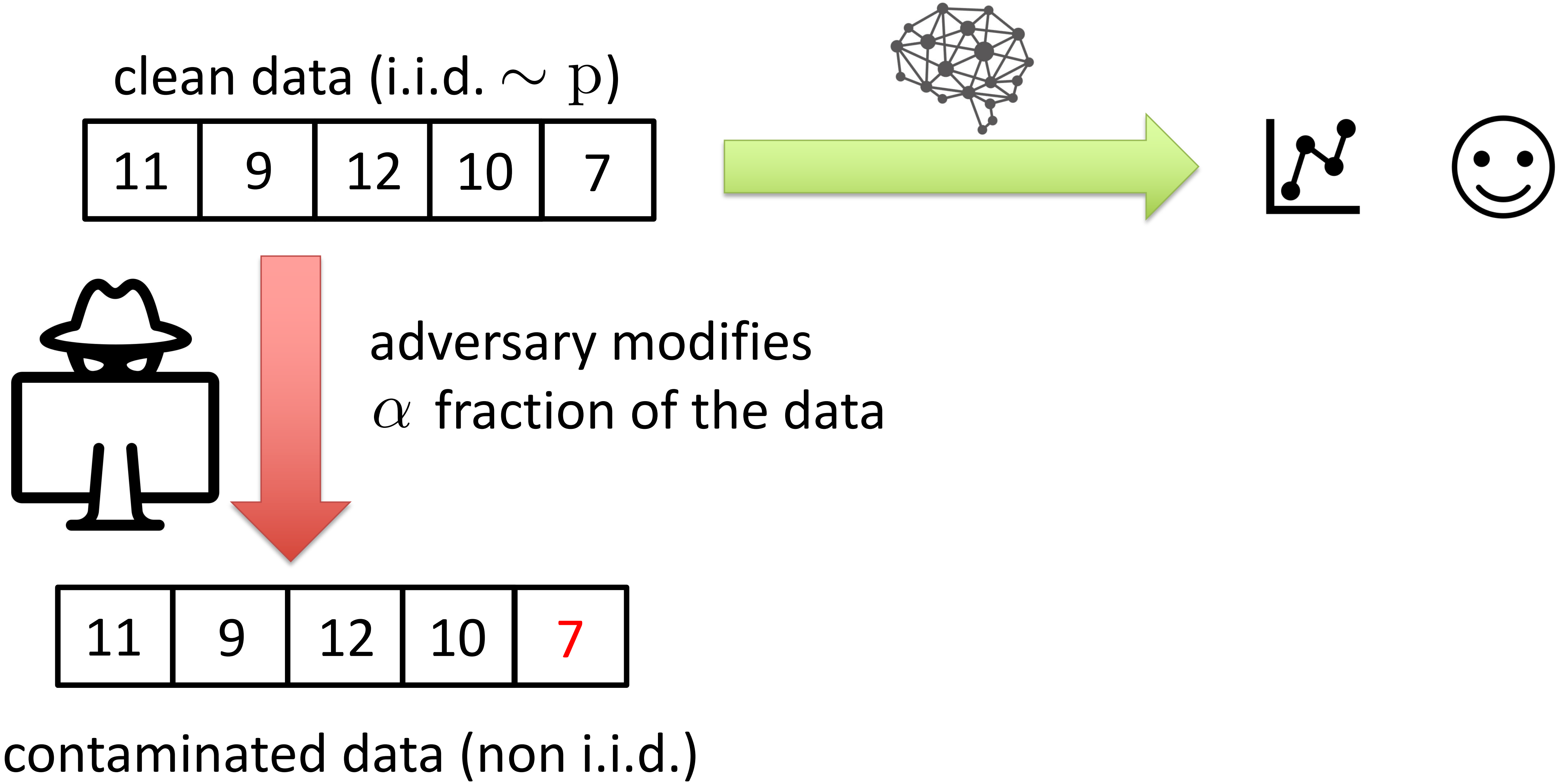
Contaminated data can lead to erroneous decisions

clean data (i.i.d. $\sim \mathrm{p}$)

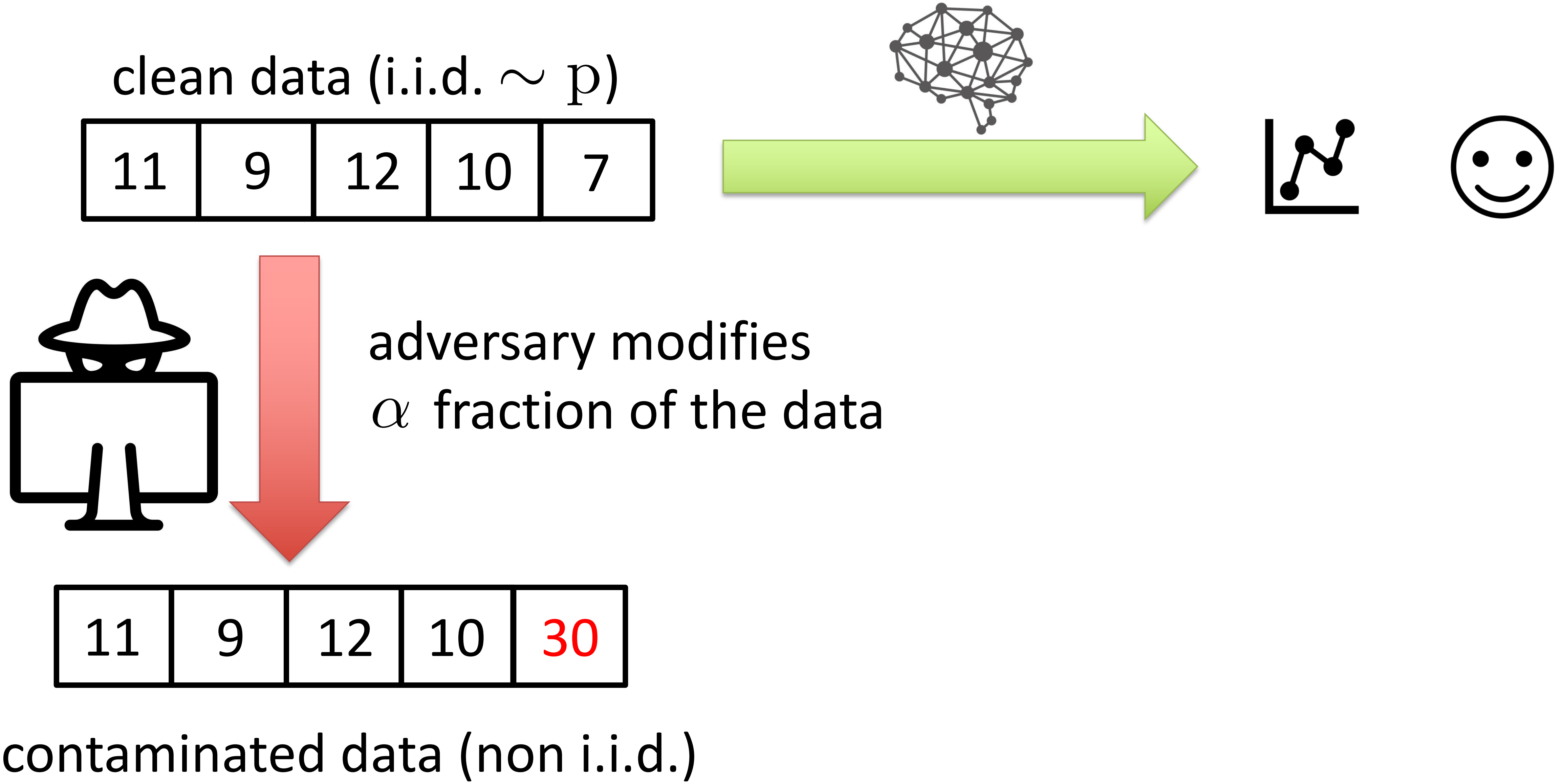
11	9	12	10	7



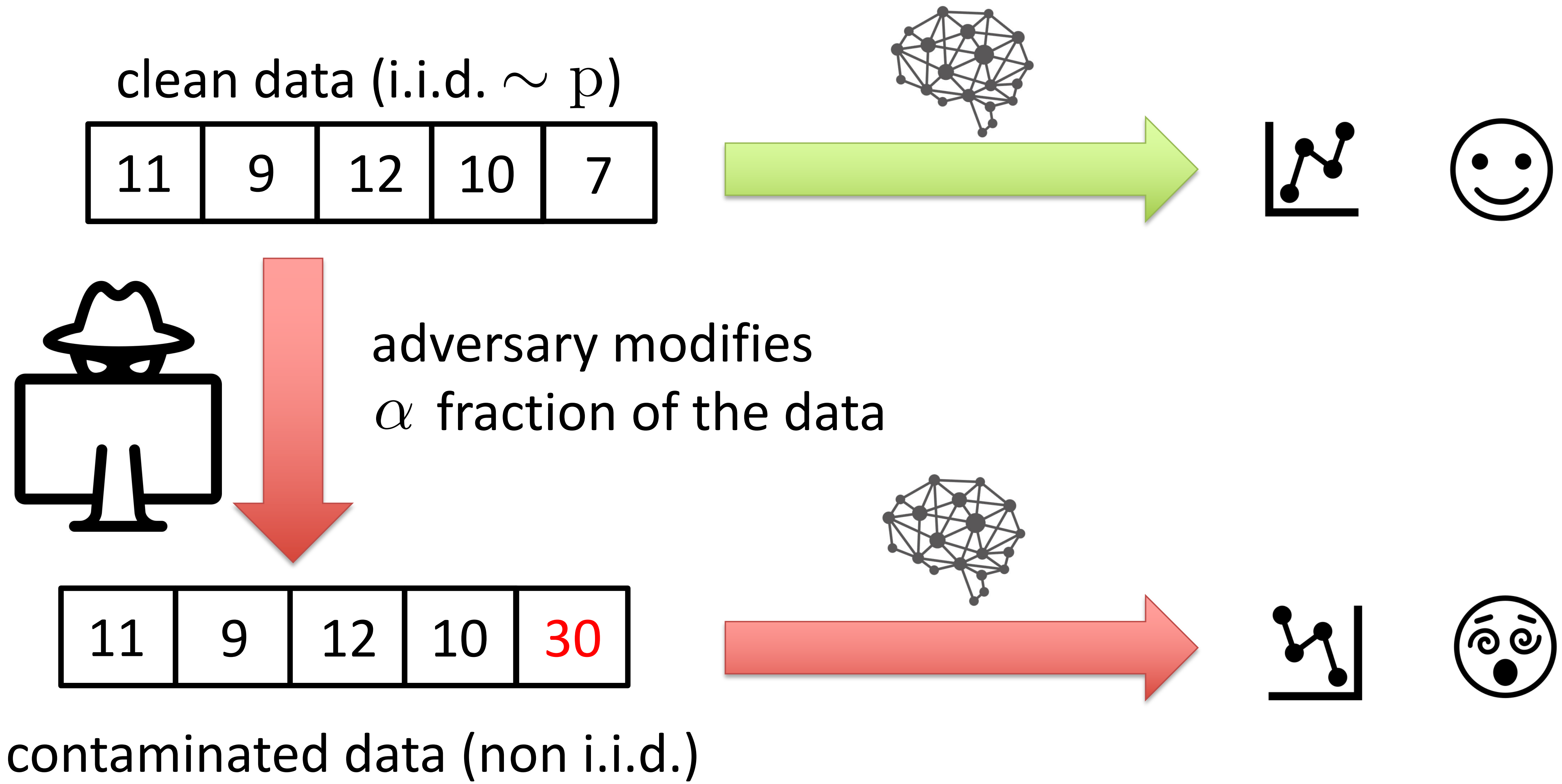
Contaminated data can lead to erroneous decisions



Contaminated data can lead to erroneous decisions



Contaminated data can lead to erroneous decisions



Mean estimation under adversarial contamination

Mean of the clean distribution

Class of distributions

$$|\widehat{\mu} - \mu_{\mathrm{p}}| \leq \sigma_{\mathrm{p}} \varepsilon(\alpha) \quad \text{with high probability} \quad \forall \mathrm{p} \in \mathcal{P}$$

Estimator of the mean

Asymptotic bias (the cost of contamination)

$$\varepsilon(\alpha) \not\to 0 \quad \text{when} \quad n \to +\infty$$

Mean estimation under adversarial contamination

Mean of the clean distribution

Class of distributions

$$|\widehat{\mu} - \boxed{\mu_{\mathrm{p}}}| \leq \sigma_{\mathrm{p}} \varepsilon(\alpha) \quad \text{with high probability} \quad \forall \mathrm{p} \in \mathcal{P}$$

Estimator of the mean

Asymptotic bias (the cost of contamination)

$$\varepsilon(\alpha) \not\to 0 \quad \text{when} \quad n \to +\infty$$

Mean estimation under adversarial contamination

Mean of the clean distribution

Class of distributions

$$|\boxed{\widehat{\mu}} - \mu_{\mathrm{p}}| \leq \sigma_{\mathrm{p}}\varepsilon(\alpha) \quad \text{with high probability} \quad \forall \mathrm{p} \in \mathcal{P}$$

Estimator of the mean

Asymptotic bias (the cost of contamination)

$$\varepsilon(\alpha) \not\to 0 \ \text{ when } \ n \to +\infty$$

Mean estimation under adversarial contamination

Mean of the clean distribution

Class of distributions

$$|\widehat{\mu} - \mu_{\mathrm{p}}| \leq \sigma_{\mathrm{p}} \boxed{\varepsilon(\alpha)} \quad \text{with high probability} \quad \forall \mathrm{p} \in \mathcal{P}$$

Estimator of the mean

Asymptotic bias (the cost of contamination)

$$\varepsilon(\alpha) \not\to 0 \quad \text{when} \quad n \to +\infty$$

Mean estimation under adversarial contamination

Mean of the clean distribution

Class of distributions

$$|\widehat{\mu} - \mu_{\mathrm{p}}| \leq \sigma_{\mathrm{p}} \varepsilon(\alpha) \quad \text{with high probability} \quad \forall \mathrm{p} \in \boxed{\mathcal{P}}$$

Estimator of the mean

Asymptotic bias (the cost of contamination)

$$\varepsilon(\alpha) \not\to 0 \quad \text{when} \quad n \to +\infty$$

Previous work on the Median-of-Means (MoM)

- Lacks analysis of the asymptotic bias
- Focuses only on the class of distributions with finite variance
- Considers a weaker contamination model

Contributions

	MoM [this work]	Trimmed mean	M-estimator
Heavy-tailed (finite variance)	$\sqrt{\alpha}$	$\sqrt{\alpha}$	$\sqrt{\alpha}$
Heavy-tailed (infinite variance)	$\alpha^{\frac{r}{1+r}}$		$\alpha^{\frac{r}{1+r}}$
Light-tailed (sub-exponential)	$\alpha^{2/3}$	$\alpha\sqrt{\log(1/\alpha)}$	

MoM is optimal for heavy-tailed distributions

Theorem. Let $\widehat{\mu}_{\mathrm{MoM}}$ be MoM with a number of blocks $k = \mathcal{O}(\alpha n)$

Minimax optimal!

- If $\mathrm{p} \in \mathcal{P}_2$, i.e., finite variance

  $$|\widehat{\mu}_{\mathrm{MoM}} - \mu_{\mathrm{p}}| \leq \mathcal{O}(\sqrt{\alpha}) \quad \text{with high probability}$$

- If $\mathrm{p} \in \mathcal{P}_{1+r}$, i.e., finite $(1+r)$-th moment

  $$|\widehat{\mu}_{\mathrm{MoM}} - \mu_{\mathrm{p}}| \leq \mathcal{O}\left(\alpha^{\frac{r}{1+r}}\right) \quad \text{with high probability}$$

MoM is sub-optimal for light-tailed distributions

Theorem. Let $\widehat{\mu}_{\mathrm{MoM}}$ be MoM with a num. of blocks $k = \mathcal{O}(\alpha^{2/3}n)$

Then, for any light-tailed distribution p

$$|\widehat{\mu}_{\mathrm{MoM}} - \mu_{\mathrm{p}}| \leq \mathcal{O}\left(\alpha^{2/3}\right)$$ with high probability.

Sub-optimal!

Theorem. There is a light-tailed distribution & attack s.t. for any k

$$|\widehat{\mu}_{\mathrm{MoM}} - \mu_{\mathrm{p}}| \geq \mathcal{O}\left(\alpha^{2/3}\right)$$ with high probability