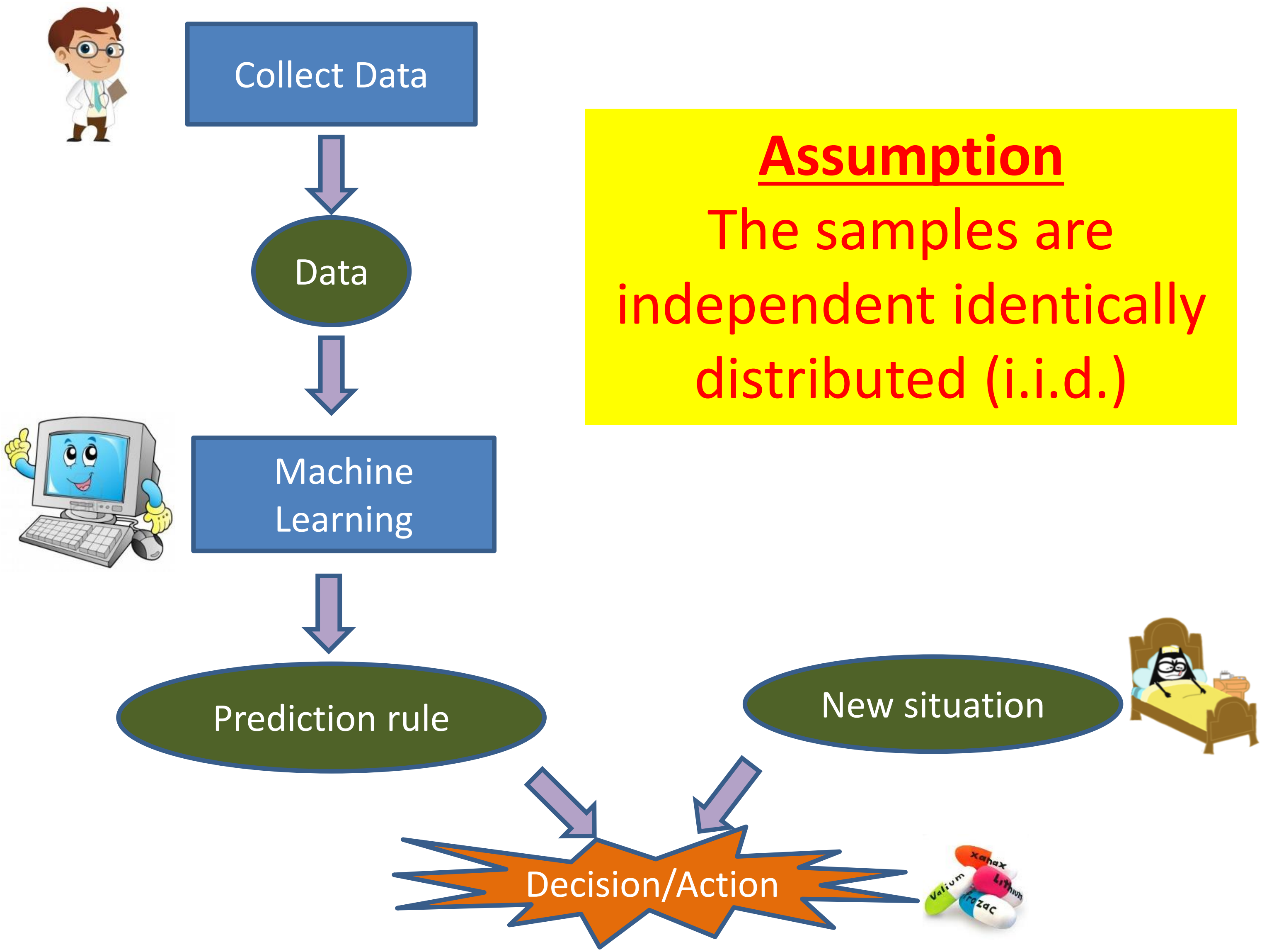
Algorithms for Easy and Hard Online Learning Environments

Yevgeny Seldin

University of Copenhagen

“Classical” (Batch) Machine Learning

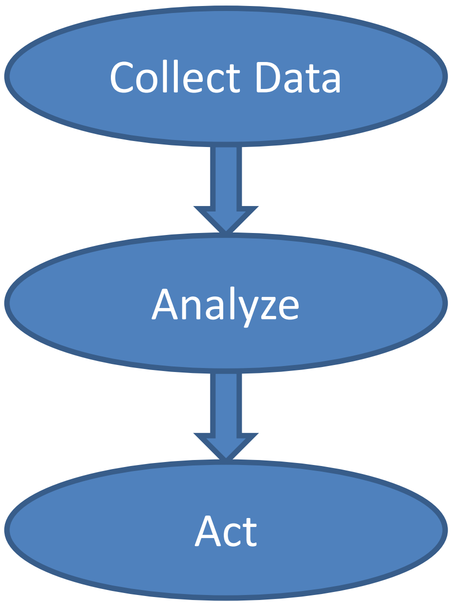


Assumption

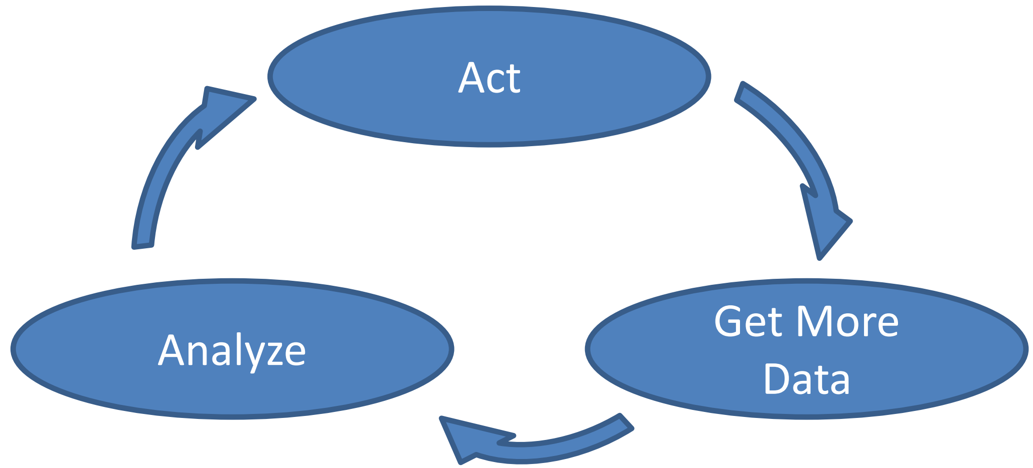
The samples are independent identically distributed (i.i.d.)

How Online is different from "batch"?

Batch Learning



Online Learning



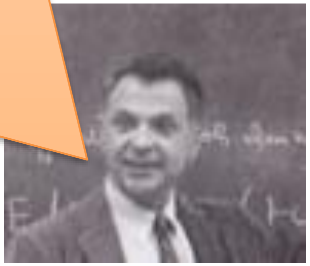
Examples

- Investment in the stock market
- Online advertising/personalization
- Online routing
- Games
- Robotics
- …

When do we need Online Learning?

Until recently, statistical theory has been restricted to the design and analysis of sampling experiments in which the size and composition of the samples are completely determined before the experimentation begins. The reasons for this are partly historical, dating back to the time when the statistician was consulted, if at all, only after the experiment was over, and partly intrinsic in the mathematical difficulty of working with anything but a fixed number of independent random variables. A major advance now appears to be in the making with the creation of a theory of the sequential design of experiments, in which the size and composition of the samples are not fixed in advance but are functions of the observations themselves.

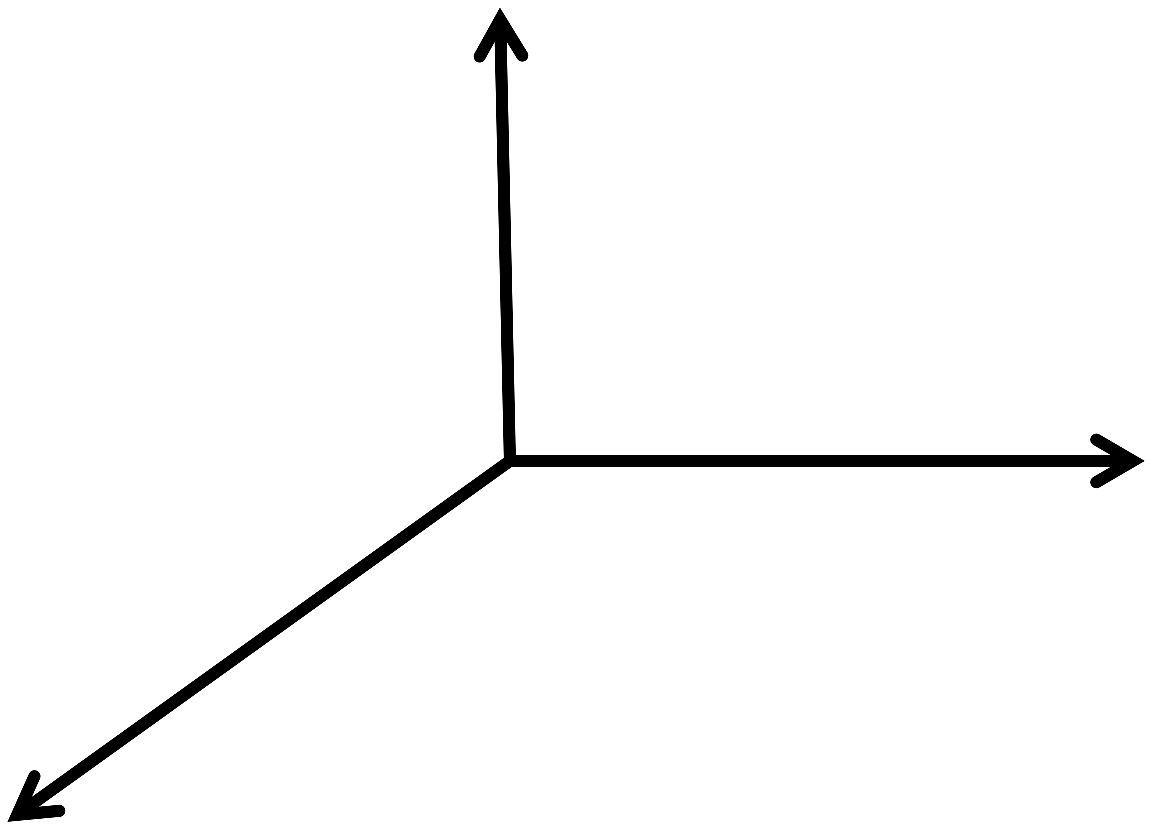
(Robbins, 1952)



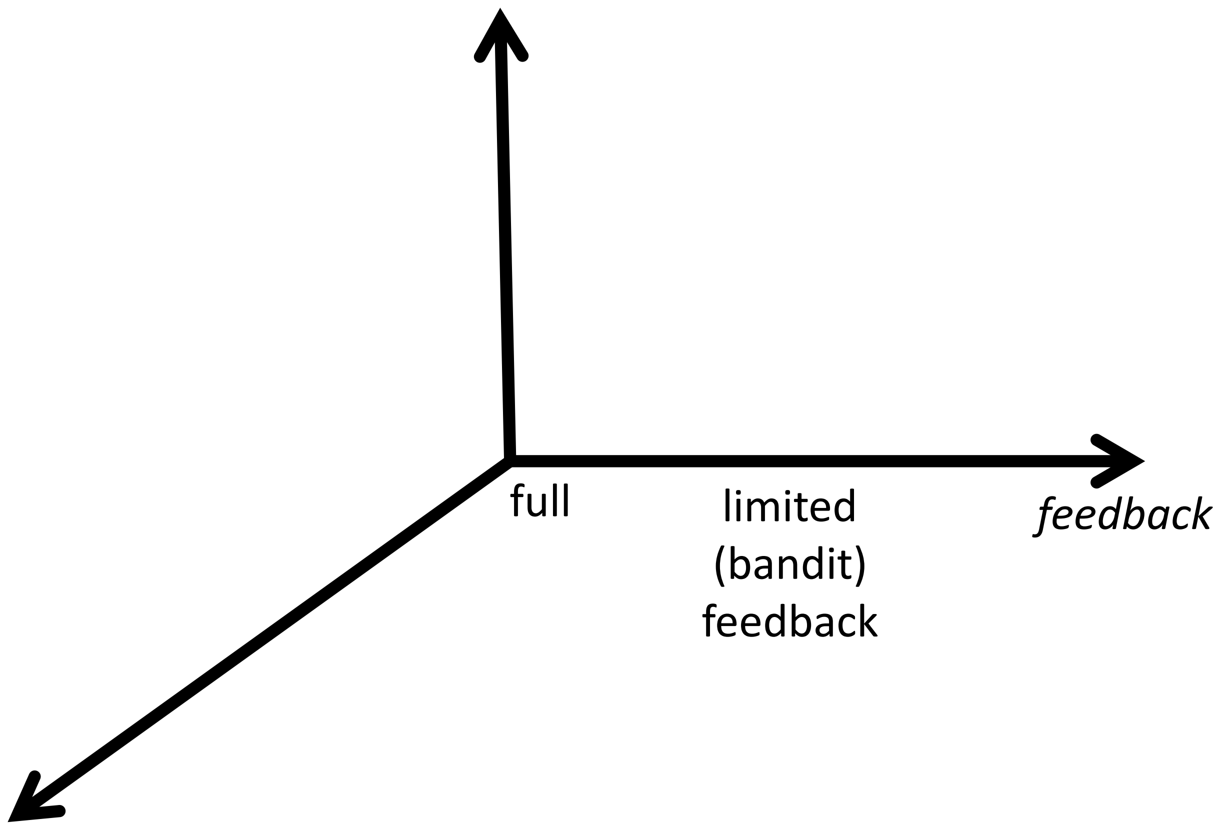
When do we need Online Learning?

- Interactive learning
- “Adversarial” game-theoretic settings
  - No i.i.d. assumption
- Intelligent data collection / experiment design
- Large-scale data analysis

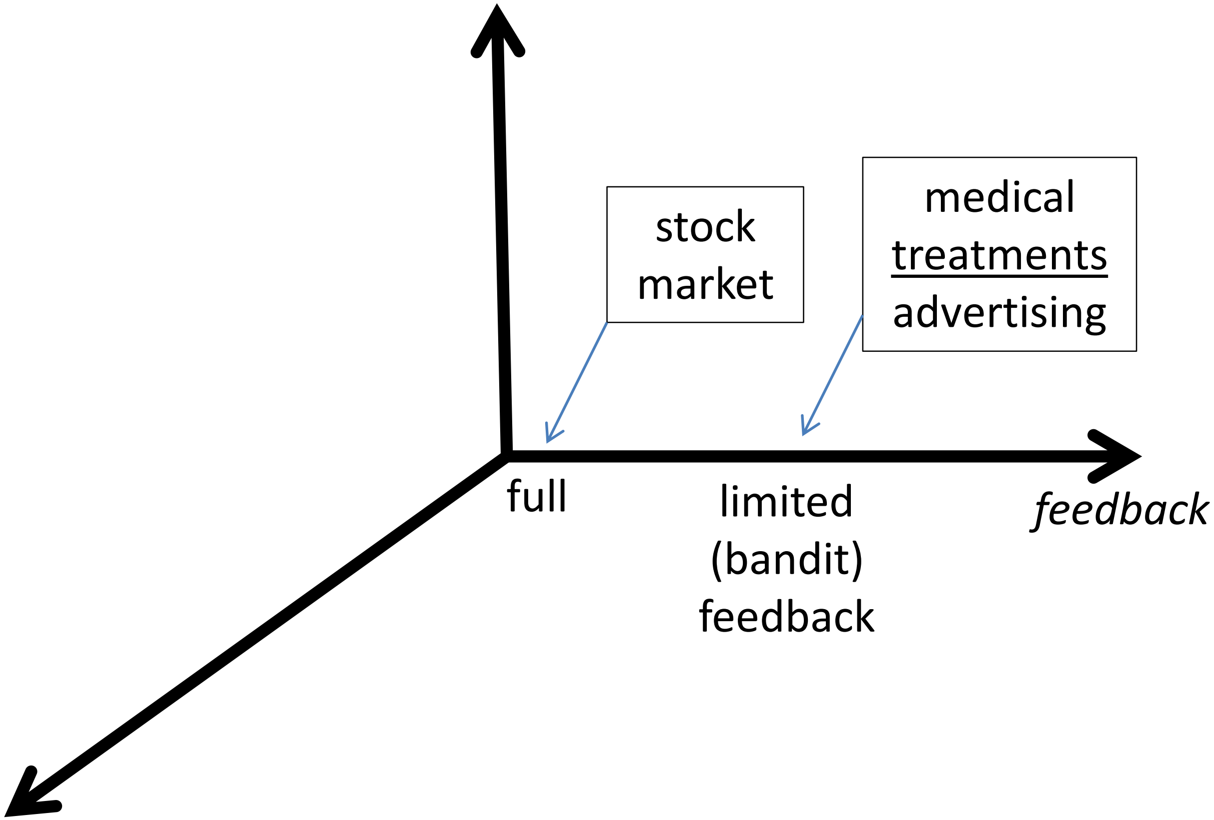
The Space of Online Learning Problems



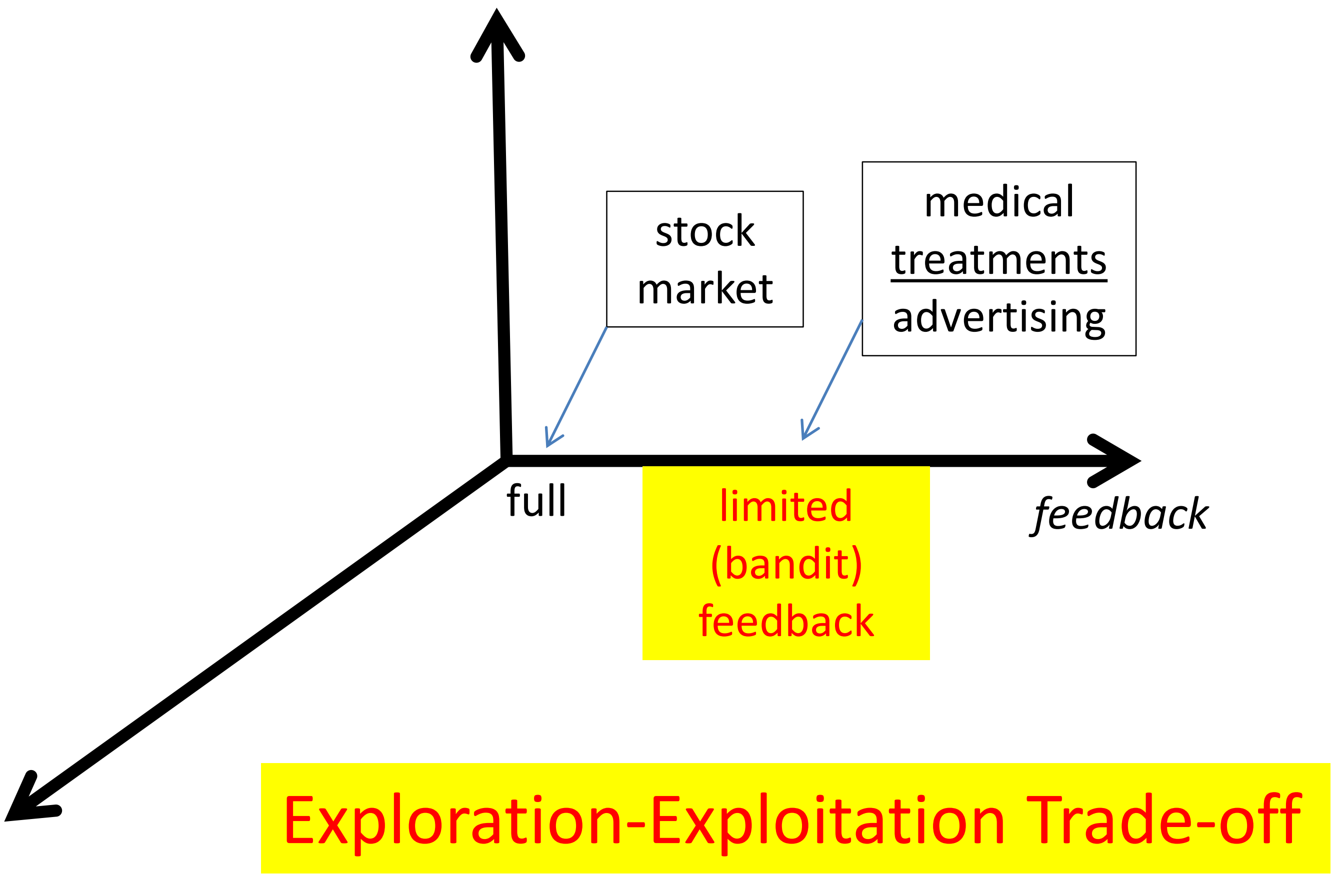
The Space of Online Learning Problems



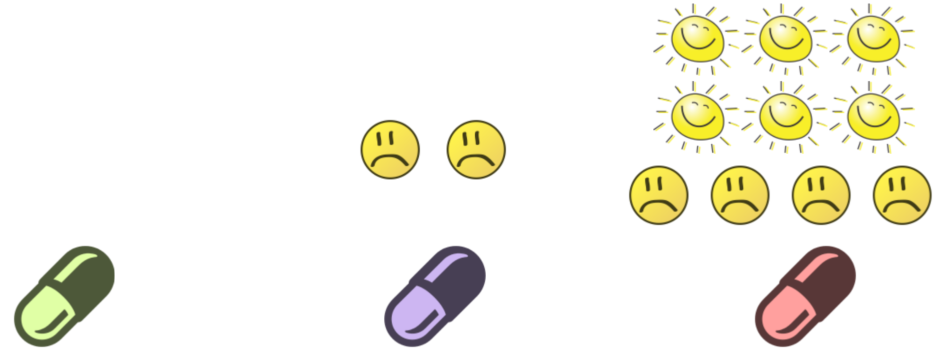
The Space of Online Learning Problems



The Space of Online Learning Problems



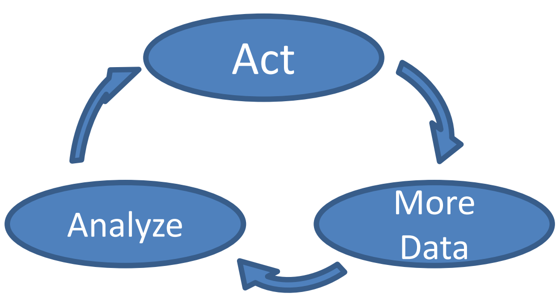
Exploration-Exploitation Trade-off



Never tried | 0/2 | 6/10

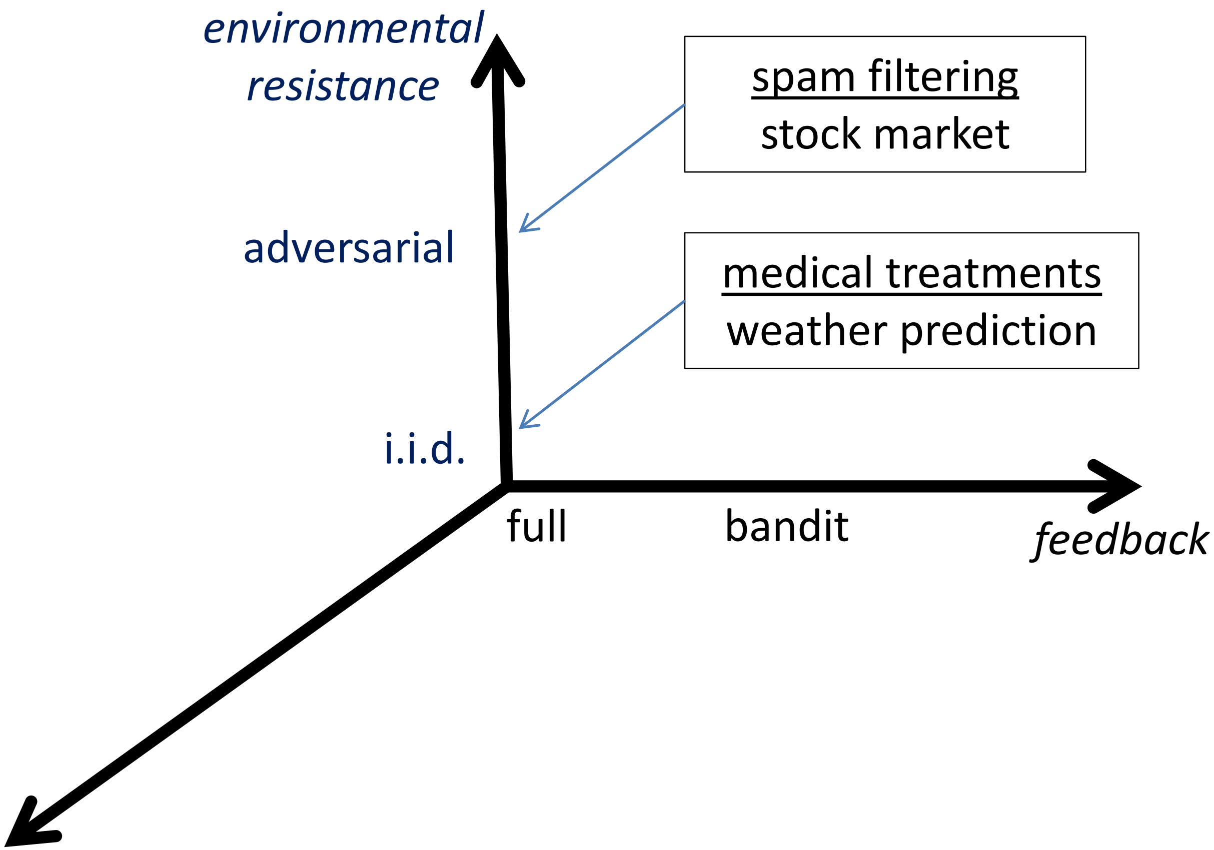
What drug to give to a new patient

When there are more patients to come…

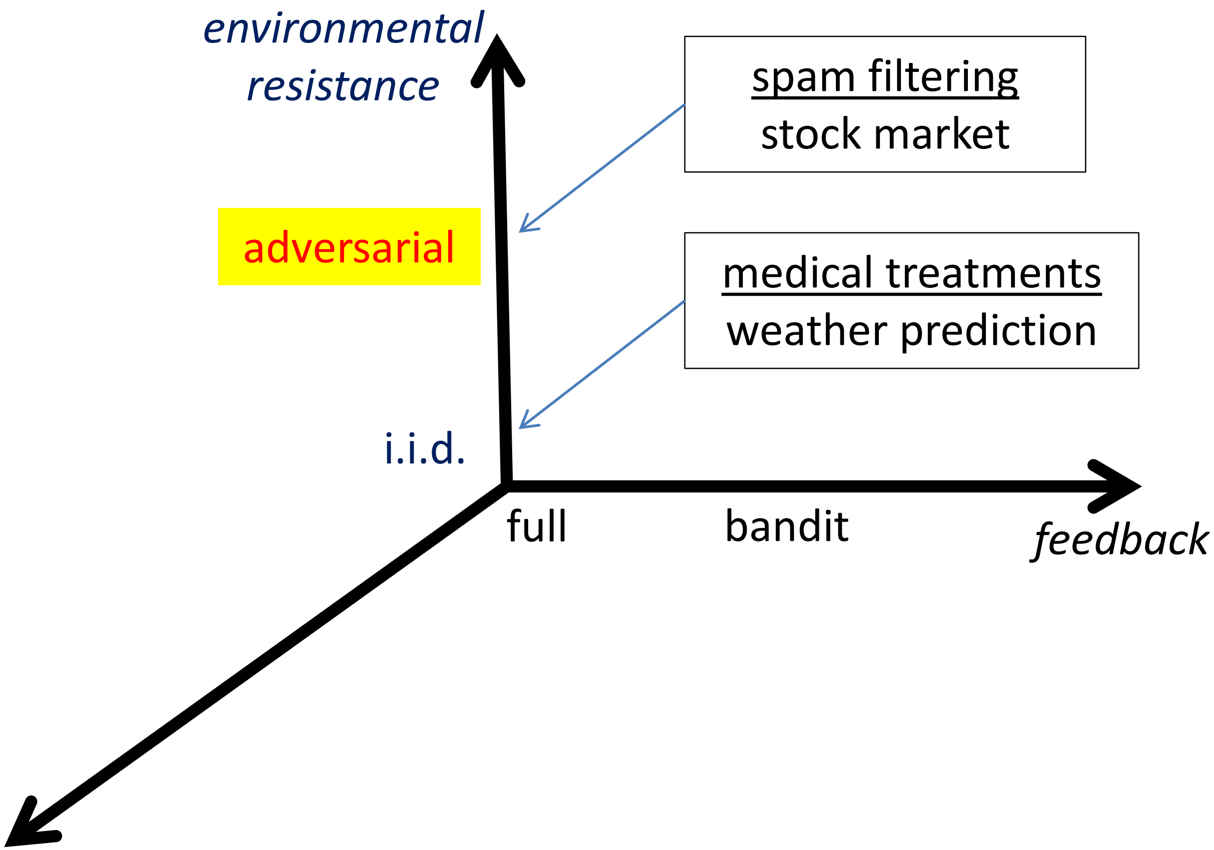


We are building the dataset for ourselves

The Space of Online Learning Problems



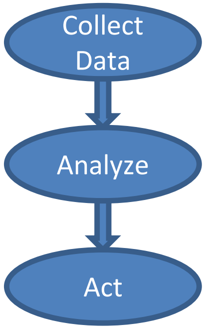
The Space of Online Learning Problems

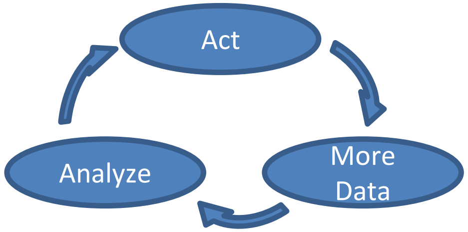


Learning in Adversarial Environments

- Game theoretic setting

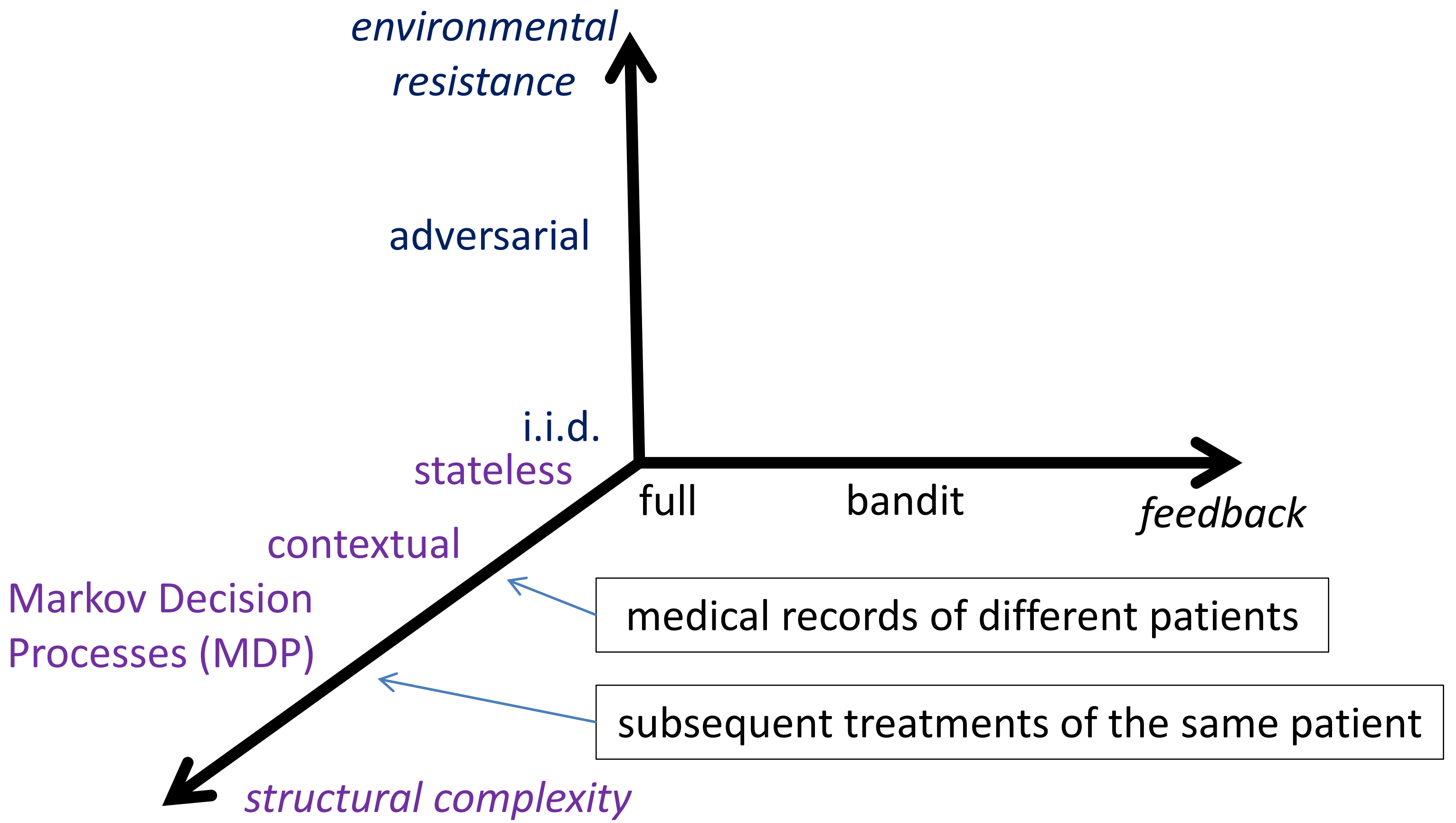
- Cannot be treated in batch learning



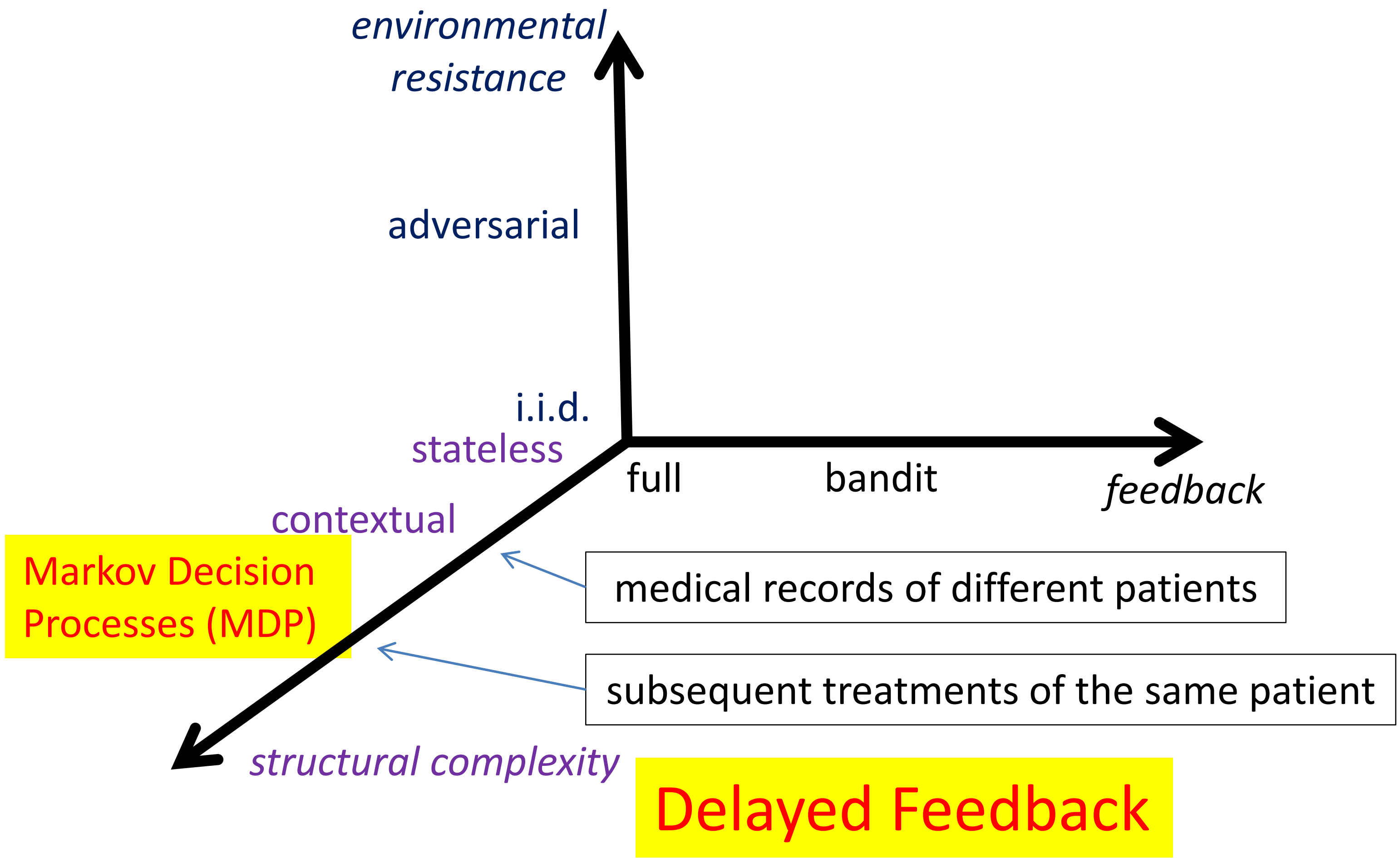


- Evaluation measure: *regret*
  - Difference in performance compared to the best choice in hindsight (out of a limited set)
  - E.g. investment revenue vs. the best stock in hindsight

The Space of Online Learning Problems



The Space of Online Learning Problems

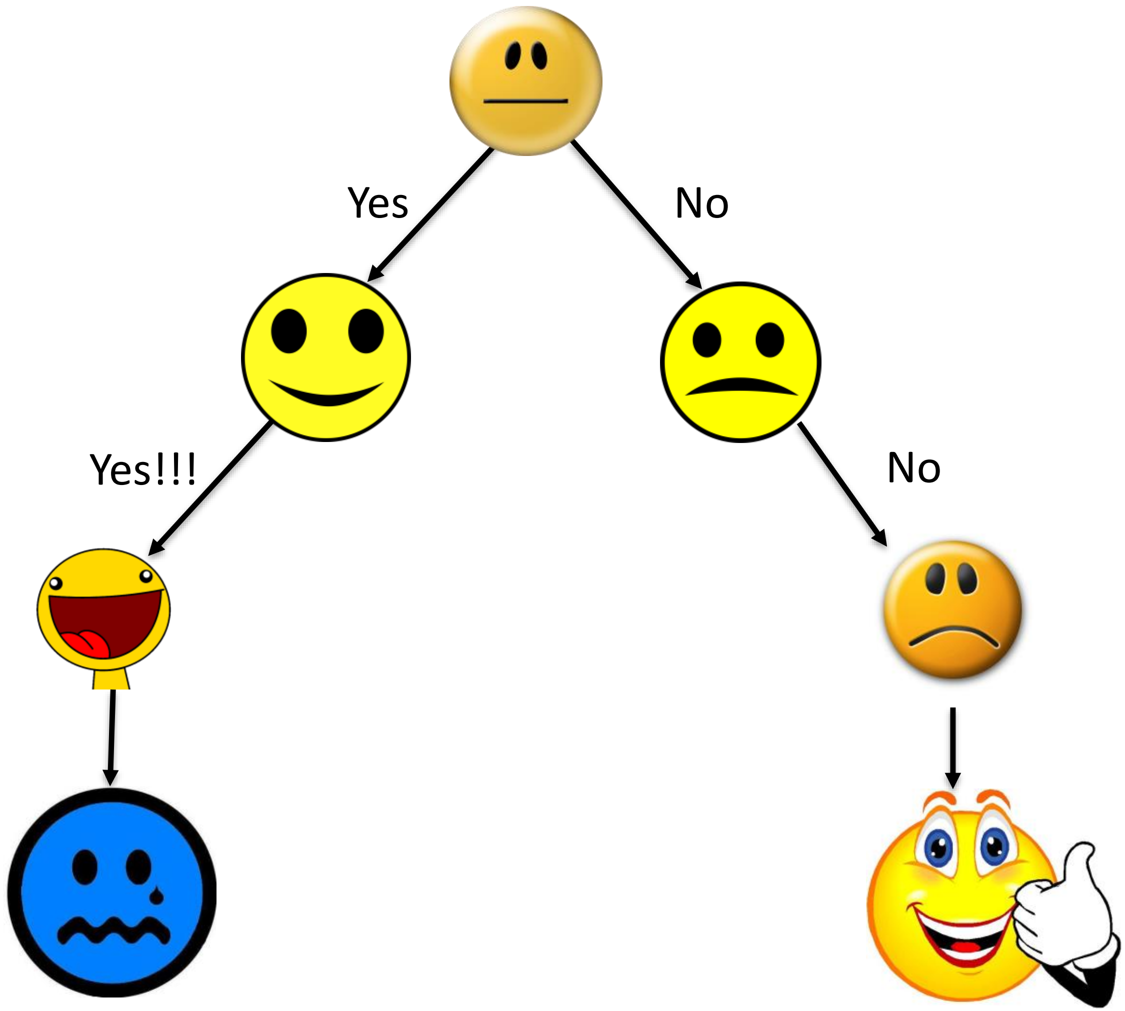


Delayed Feedback

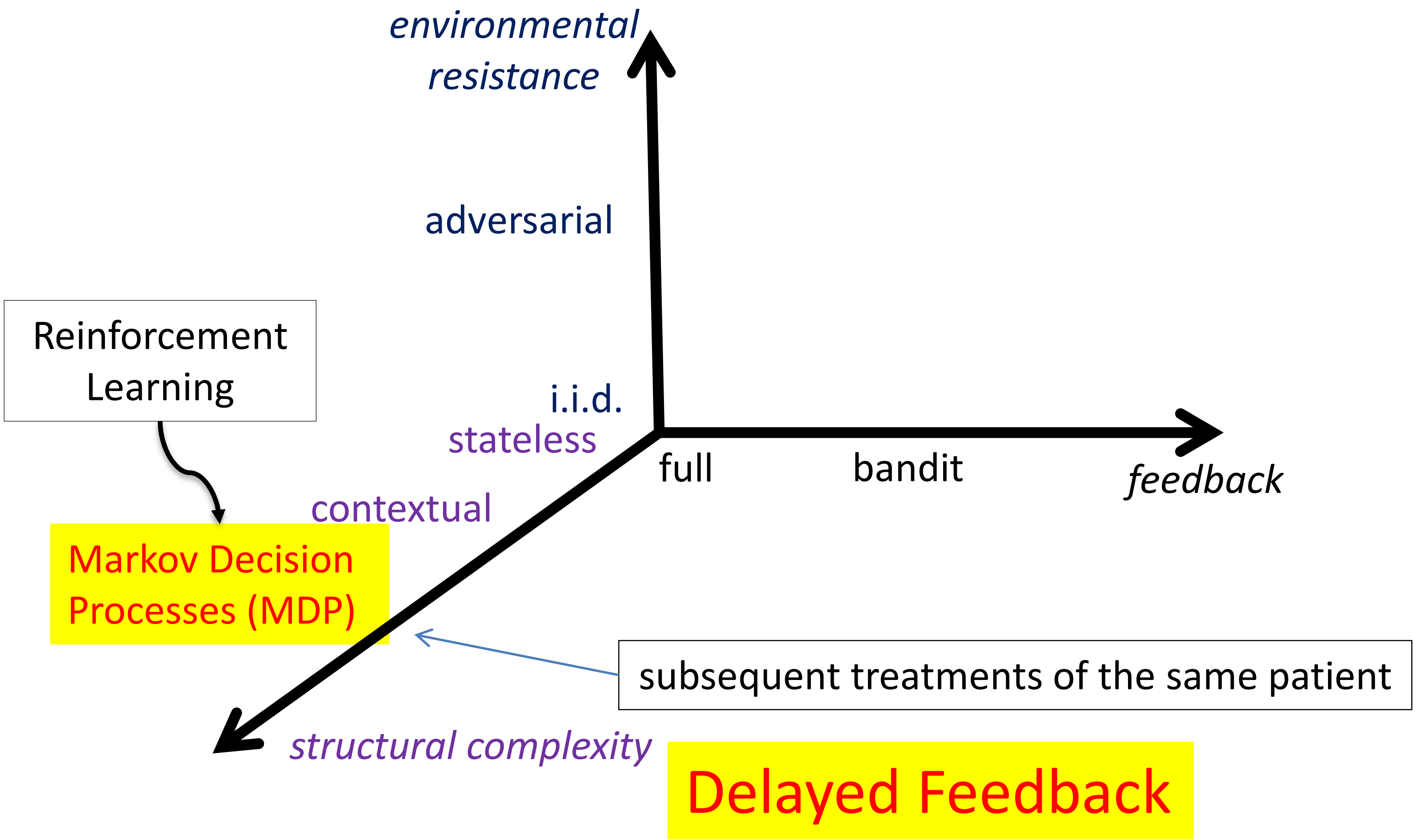
Would you like to have a drink?

Would you like to have a drink?

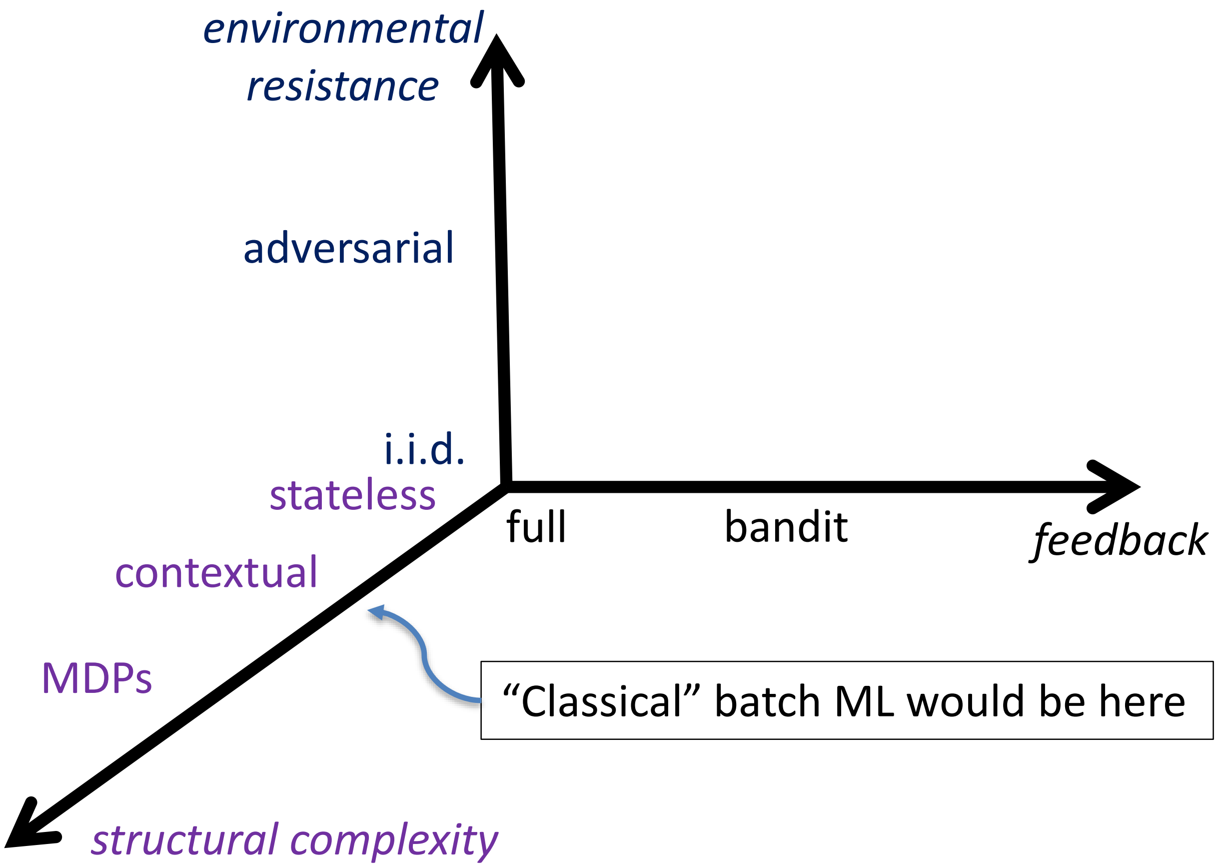
The next morning….



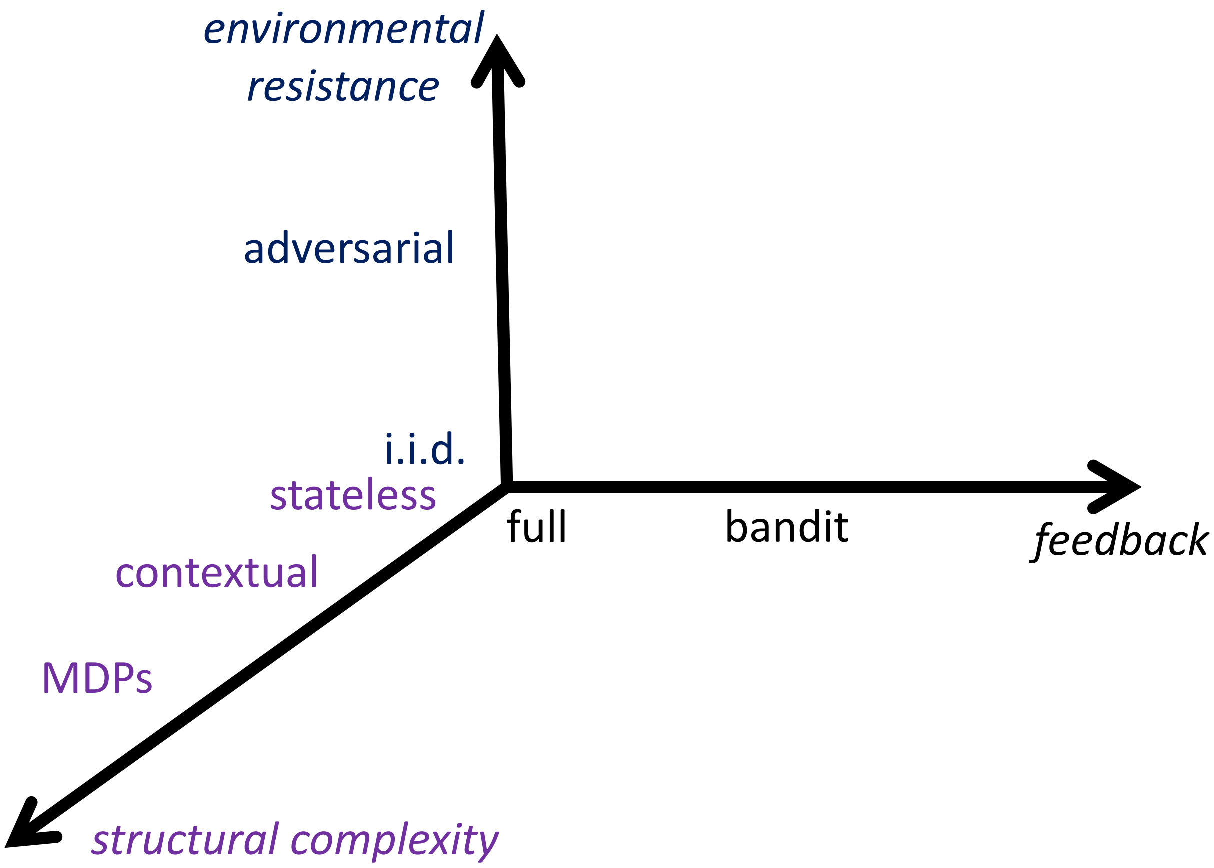
The Space of Online Learning Problems



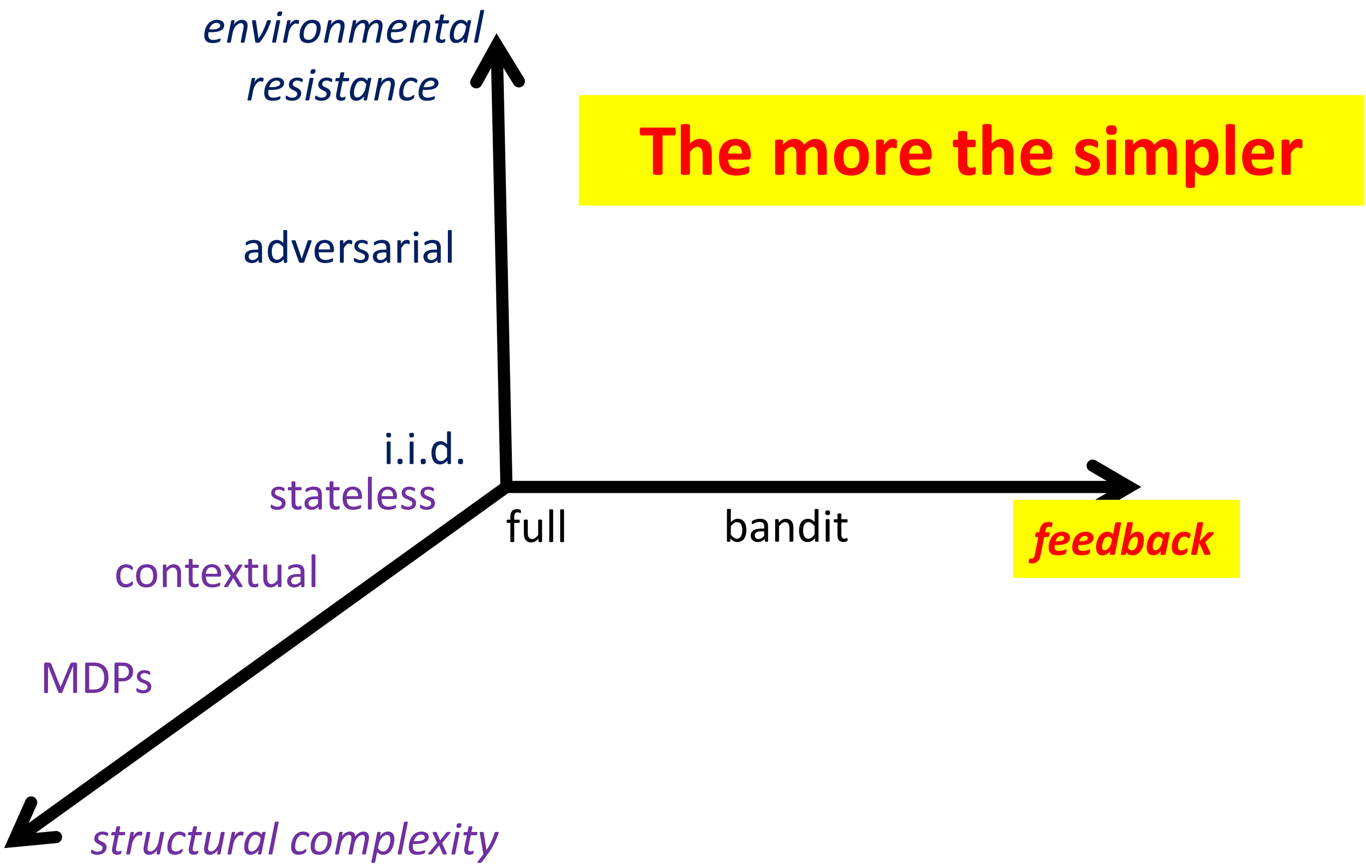
The Space of Online Learning Problems



Simplicities along the axes

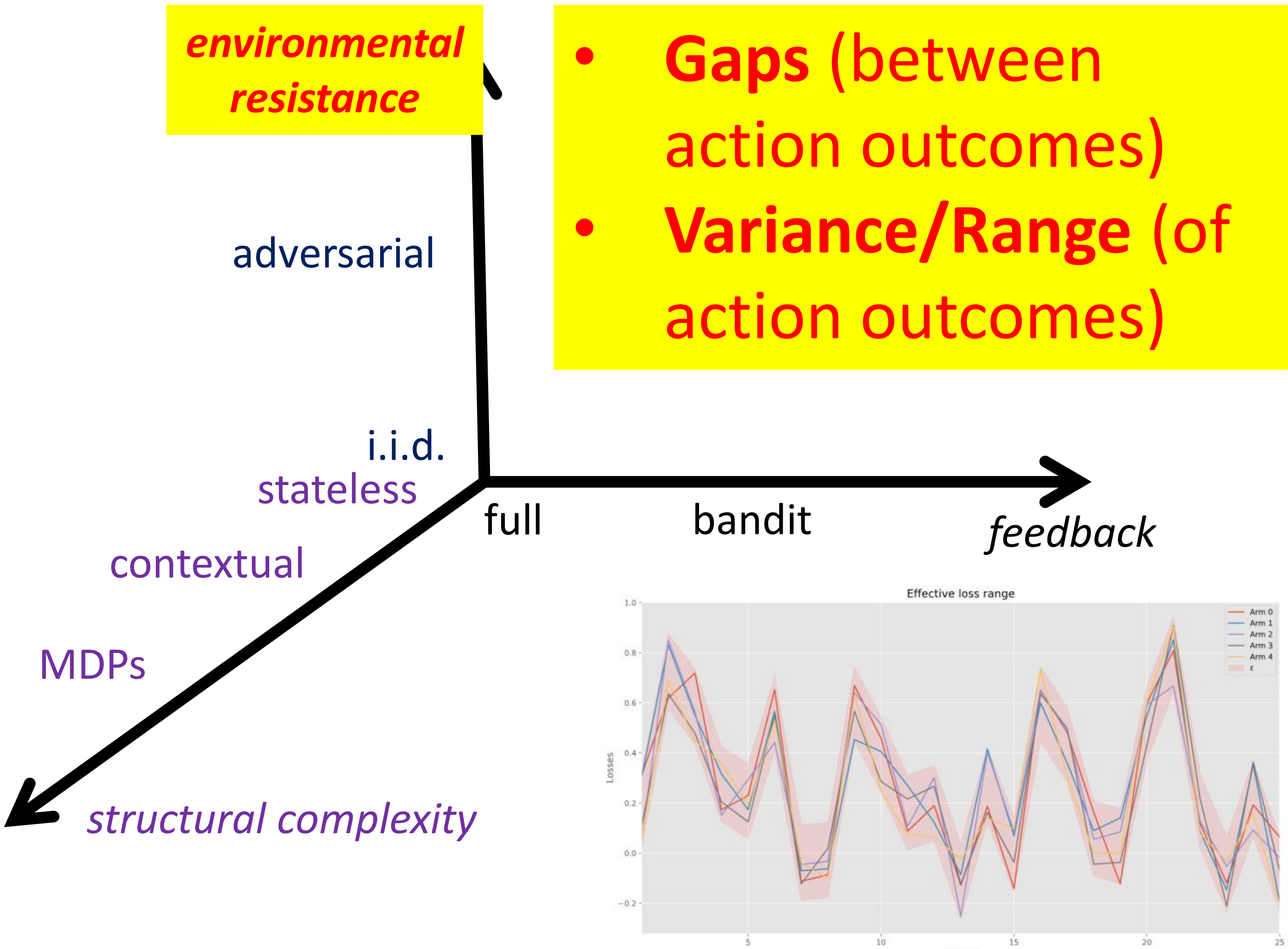


Simplicities along the axes

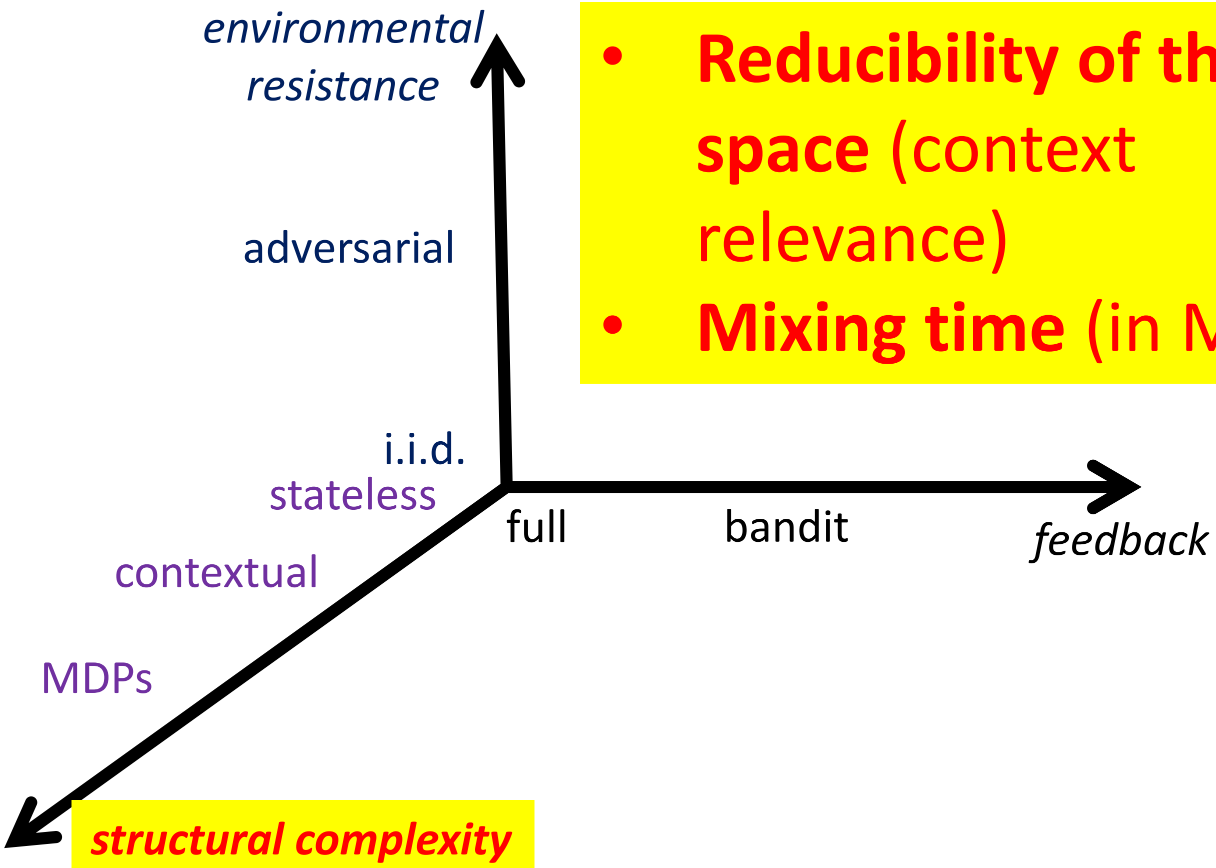


Simplicities along the axes

- **Gaps** (between action outcomes)
- **Variance/Range** (of action outcomes)

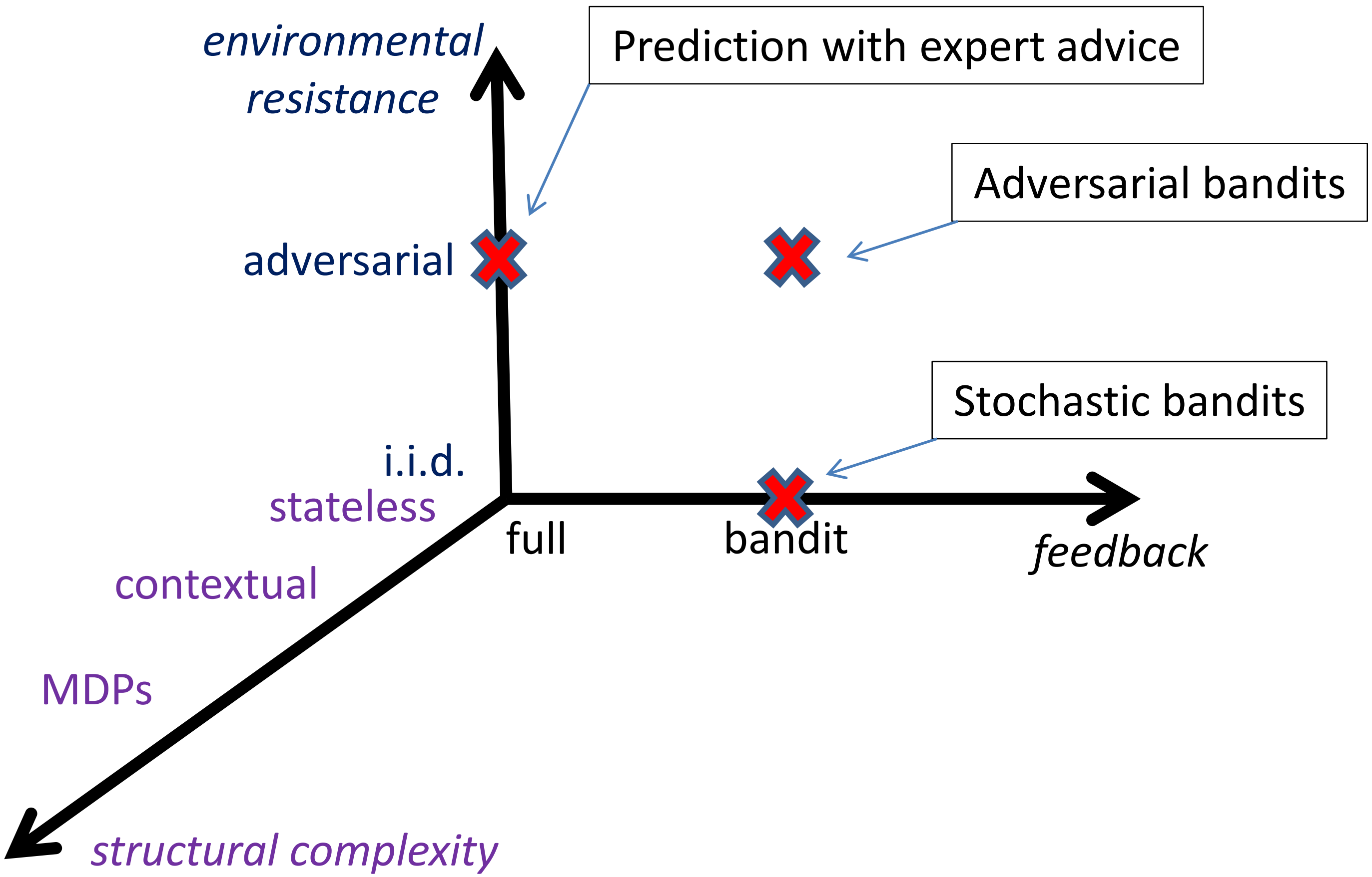


Simplicities along the axes

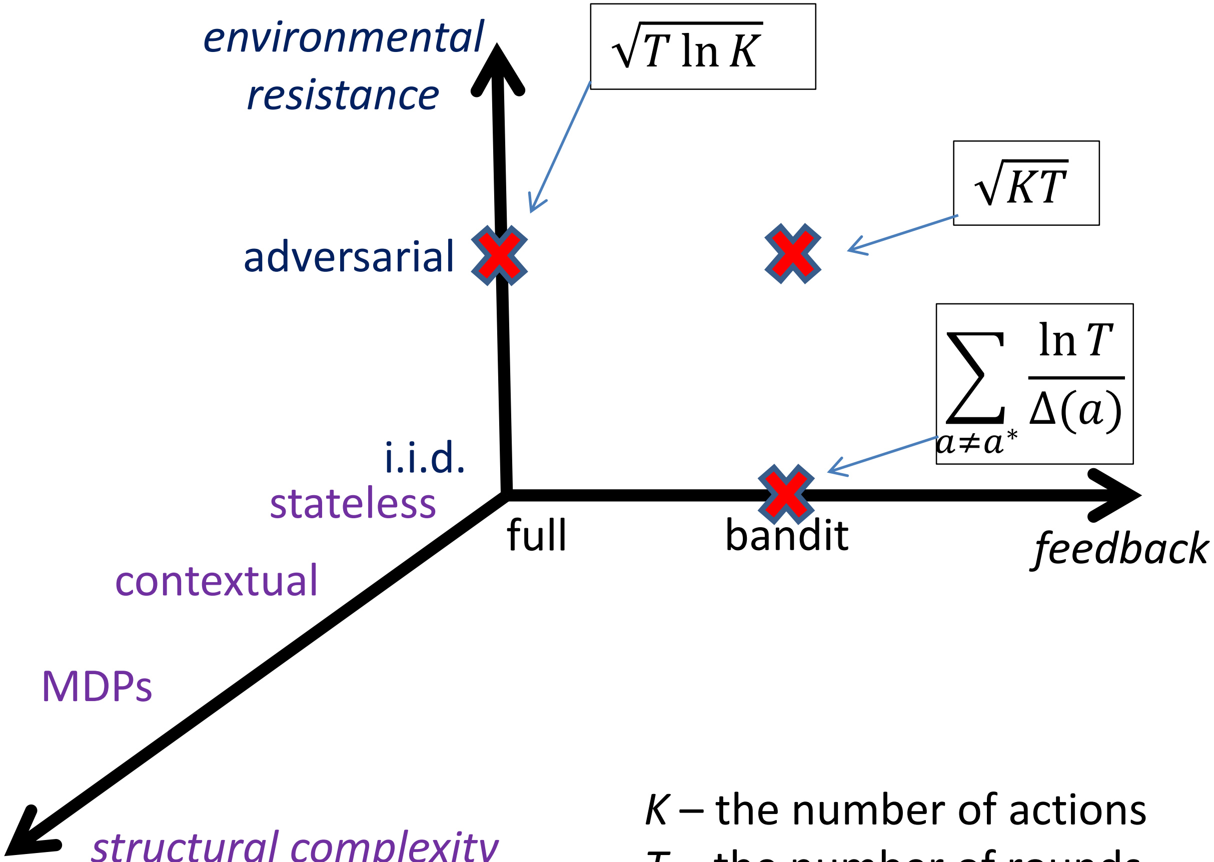


- **Reducibility of the state space** (context relevance)
- **Mixing time** (in MDPs)

Some “standard” algorithms

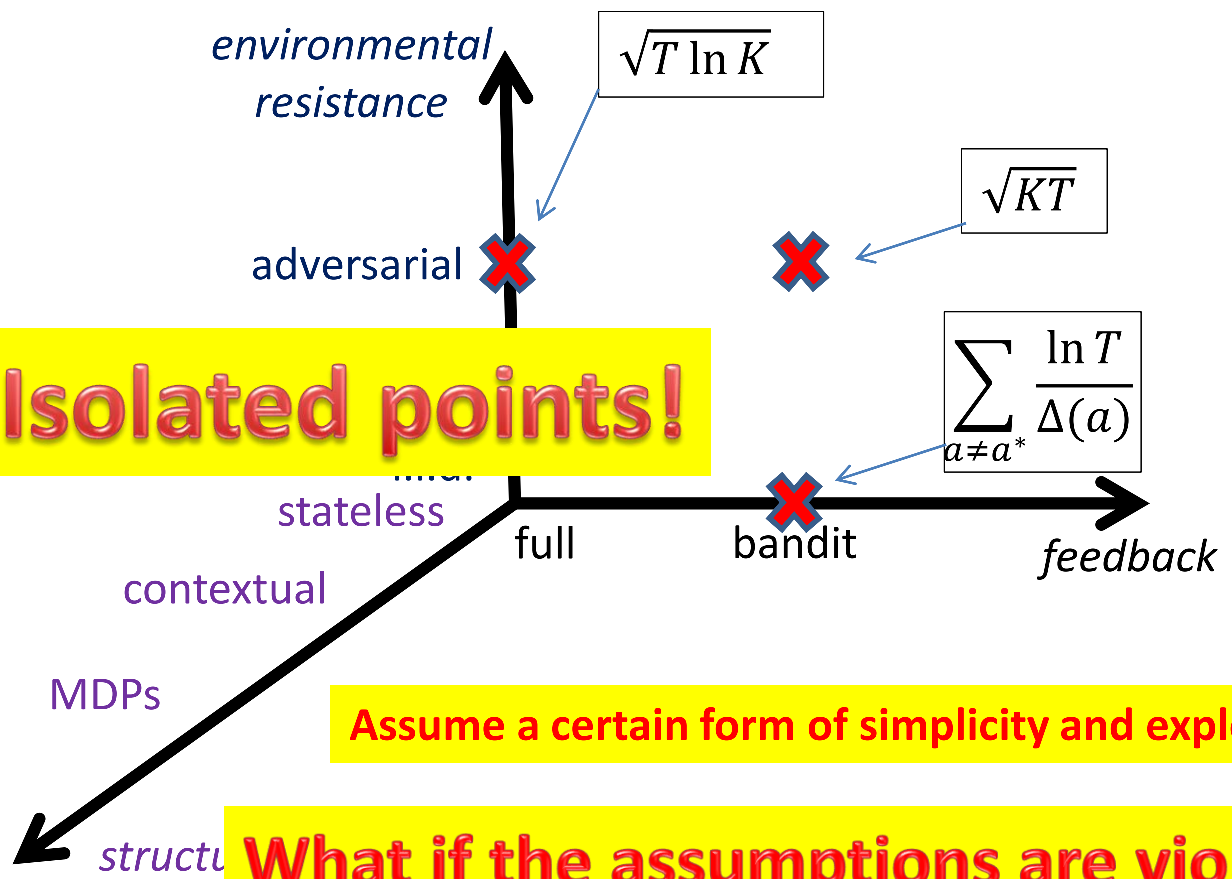


Typical performance scaling



K – the number of actions
T – the number of rounds
$\Delta(a)$ – suboptimality gap

“Standard” algorithms

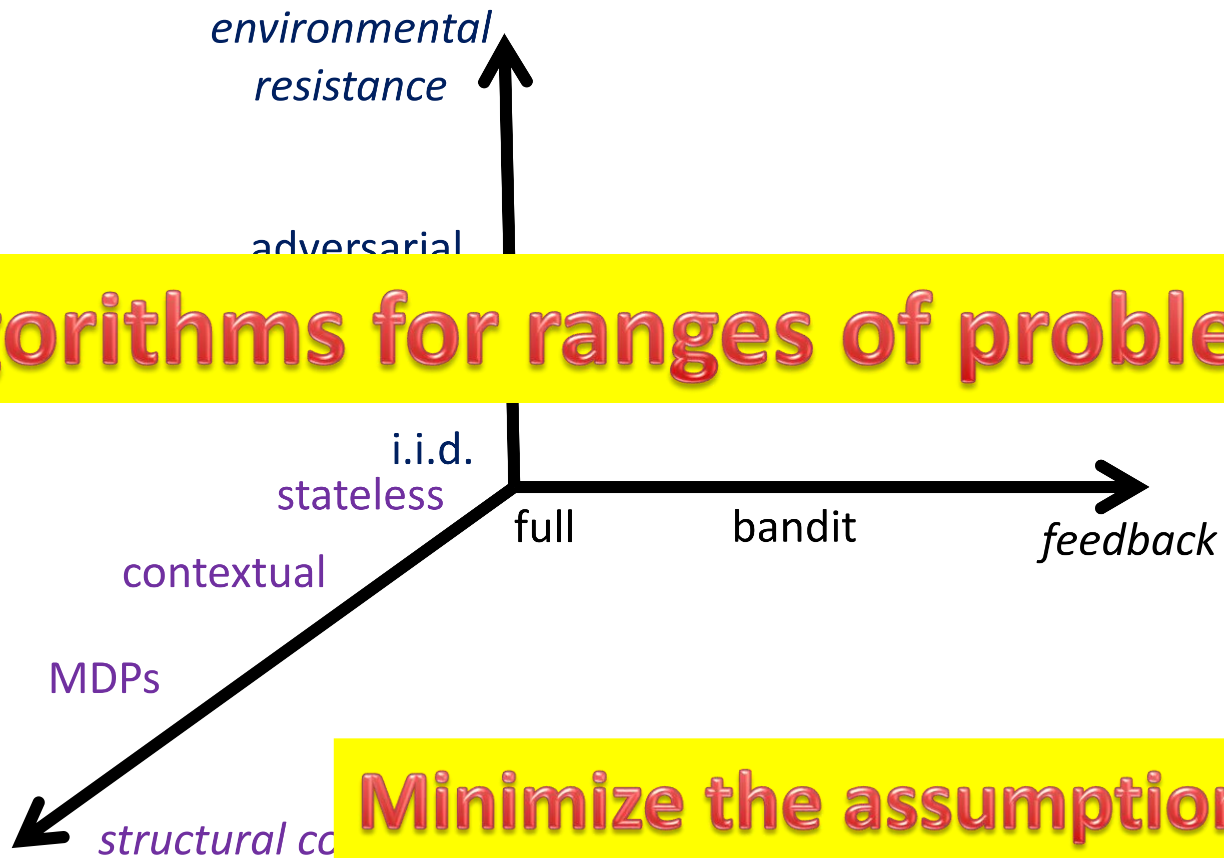


Isolated points!

Assume a certain form of simplicity and exploit it

What if the assumptions are violated?

What are we busy with?

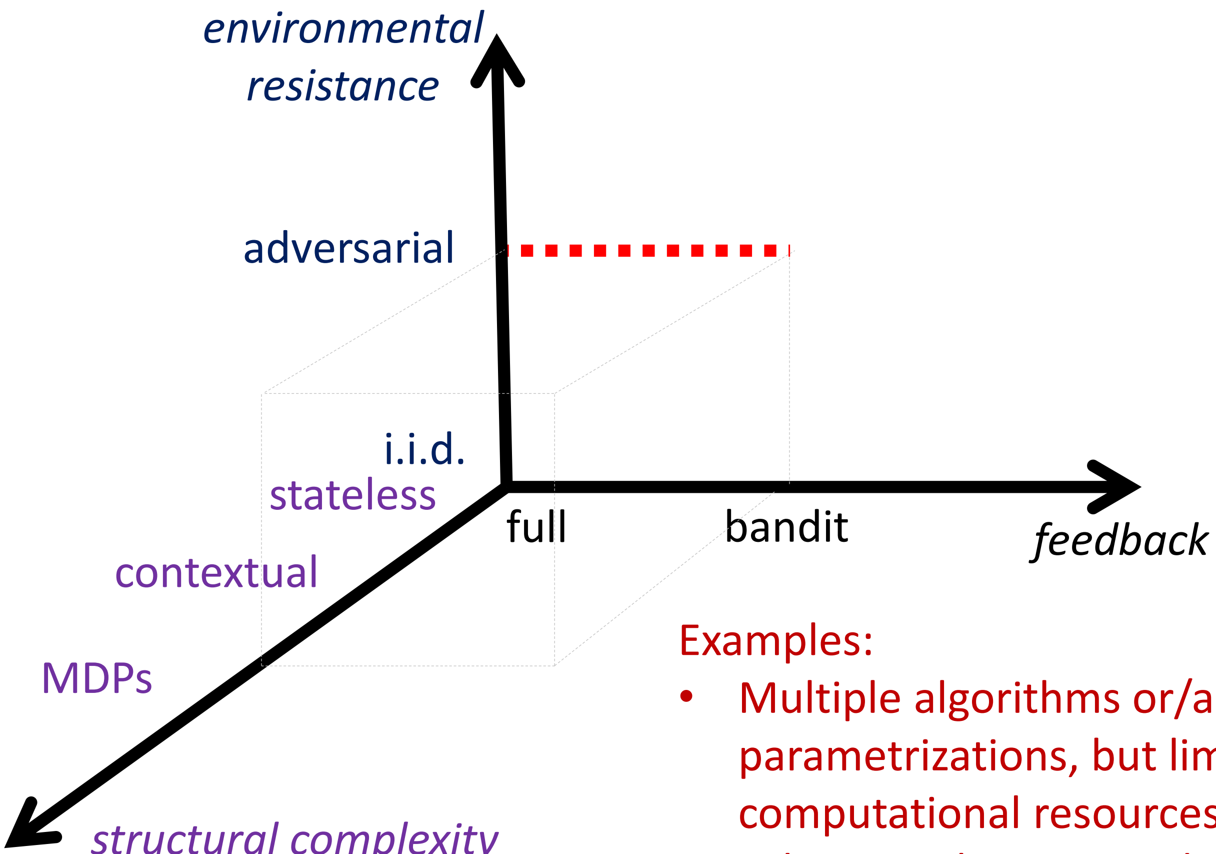


Algorithms for ranges of problems

Minimize the assumptions

Prediction with limited advice / Bandits with paid observations

[Seldin, Bartlett, Cramer, Abbasi-Yadkori, ICML, 2014]

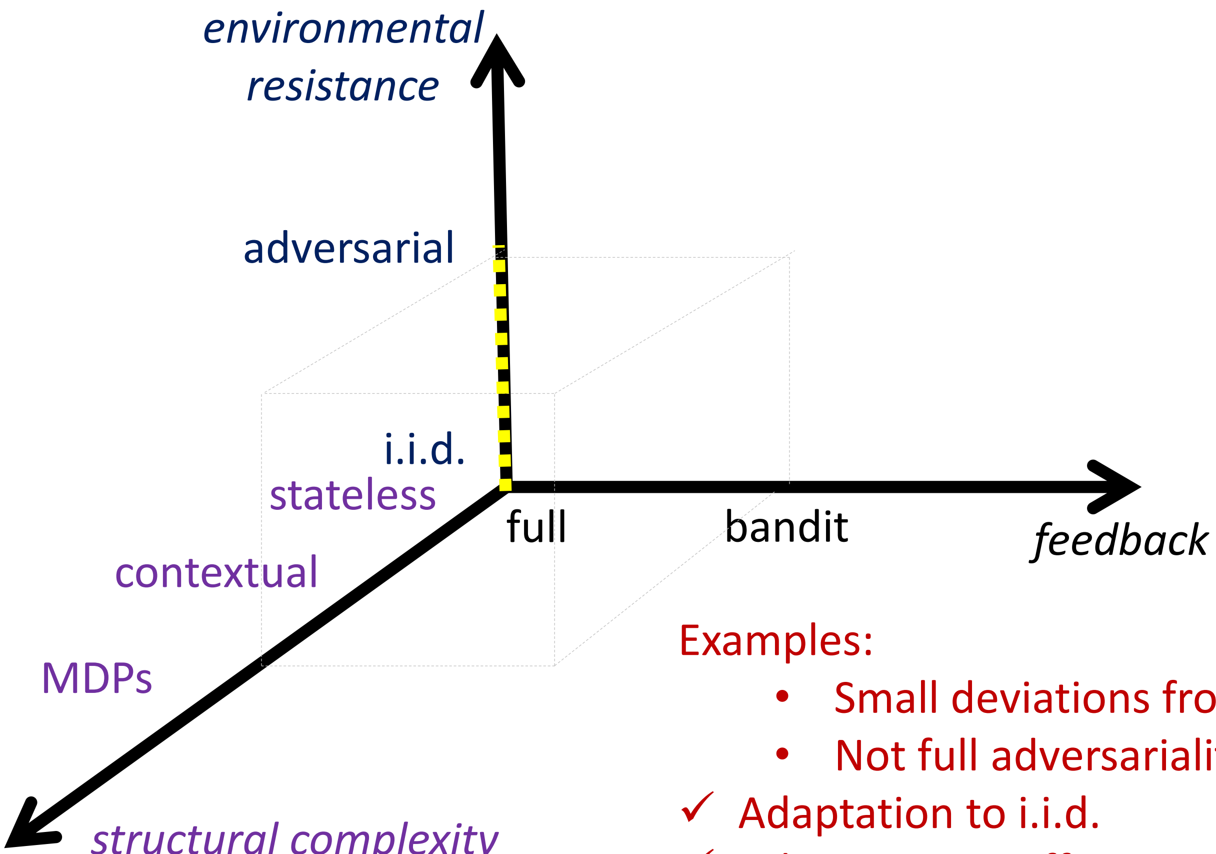


Examples:

- Multiple algorithms or/and parametrizations, but limited computational resources. Only a subset can be executed.
- Expensive resources

Environmental resistance in full info

[Koolen & van Erven, COLT, 2015, Luo & Schapire, COLT, 2015, Wintenberger, 2015, van Erven, Kotłowski & Warmuth, COLT, 2014, Gaillard, Stoltz & van Erven, COLT 2014, Cesa-Bianchi, Mansour & Stoltz, MLJ, 2007, …]



Examples:

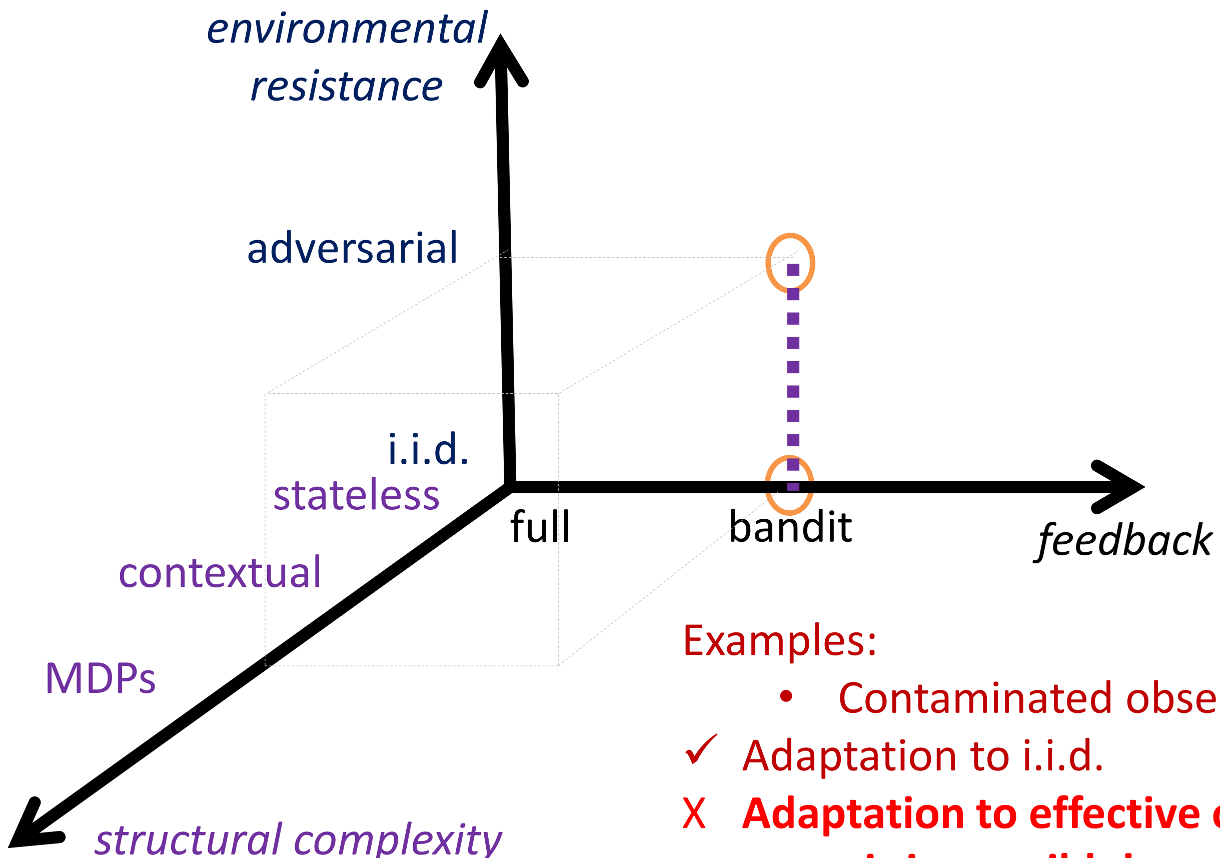
- Small deviations from i.i.d.
- Not full adversariality

✓ Adaptation to i.i.d.

✓ Adaptation to effective outcome range

Environmental resistance in bandits

[Bubeck & Slivkins, COLT, 2012, Seldin & Slivkins, ICML, 2014, Auer & Chiang, COLT, 2016, Seldin & Lugosi, COLT, 2017, Wei & Luo, COLT, 2018, **Zimmert & Seldin, AISTATS, 2019**]



Examples:

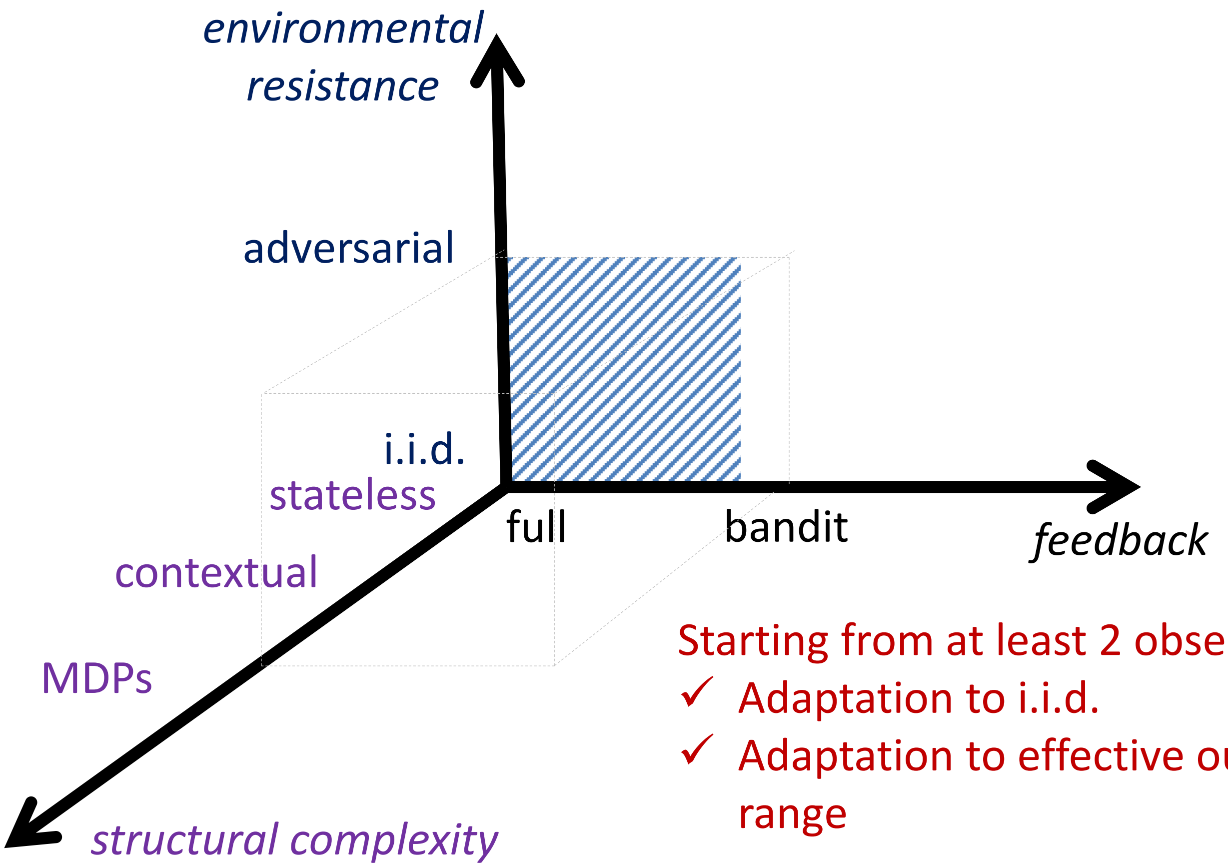
- Contaminated observations

✓ Adaptation to i.i.d.

X **Adaptation to effective outcome range is impossible! [Gershinovitz & Lattimore, NIPS, 2016]**

Prediction with Limited Advice

[Thune & Seldin, NIPS, 2018]

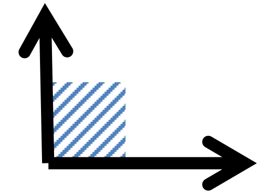


Starting from at least 2 observations

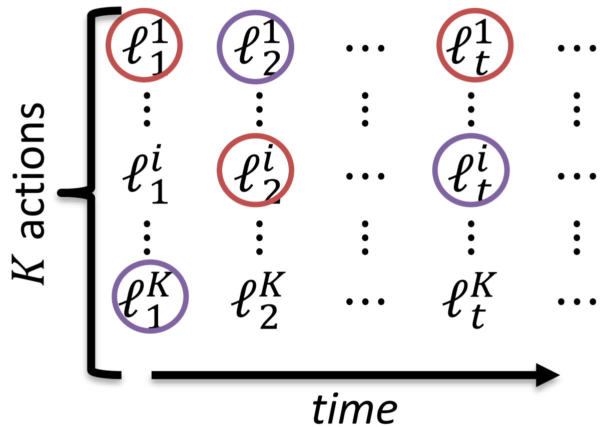
✓ Adaptation to i.i.d.

✓ Adaptation to effective outcome range

Prediction with Limited Advice



Problem Setting

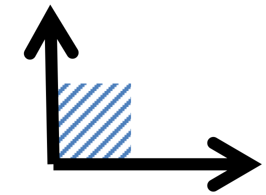


- Adversarial
  - ℓ_t^i arbitrary in [0,1]
- I.I.D.
  - $\mathbb{E}\left[\ell_t^i\right] = \mu_i$
  - Gaps: $\Delta_i = \mu_i - \min_j \mu_j$
- Effective Loss Range ε
  - $\forall i, j, t$: $\left|\ell_t^i - \ell_t^j\right| \leq \varepsilon$

- Evaluation measure: pseudo-regret
  - $R_T = \mathbb{E}\left[\sum_{t=1}^{T} \ell_t^{A_t}\right] - \min_{i \in \{1,\ldots,K\}} \mathbb{E}\left[\sum_{t=1}^{T} \ell_t^i\right]$

SODA: Second Order Difference Adjustment

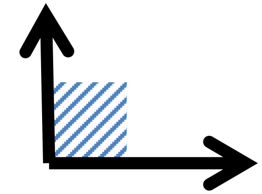
[Thune & Seldin, NIPS, 2018]



- The primary action A_t sampled according to
  - $p_t(i) \propto e^{-\eta_t D_{t-1}^i - \eta_t^2 S_{t-1}^i}$
- The secondary observation B_t sampled
  - uniformly from $\{1, \dots, K\} \setminus \{\mathrm{A}_t\}$
- Unbiased loss difference estimates
  - $\widetilde{\Delta\ell}_s^i = (K-1)\mathbb{I}(B_t = i)\left(\ell_t^{B_t} - \ell_t^{A_t}\right)$
- Loss difference estimator and its second moment
  - $D_t^i = \sum_{s=1}^t \widetilde{\Delta\ell}_s^i \qquad S_t^i = \sum_{s=1}^t \left(\widetilde{\Delta\ell}_s^i\right)^2$
- The learning rate
  - $\eta_t \approx \sqrt{\dfrac{\ln K}{\max_i S_{t-1}^i}}$

SODA: Regret Guarantees

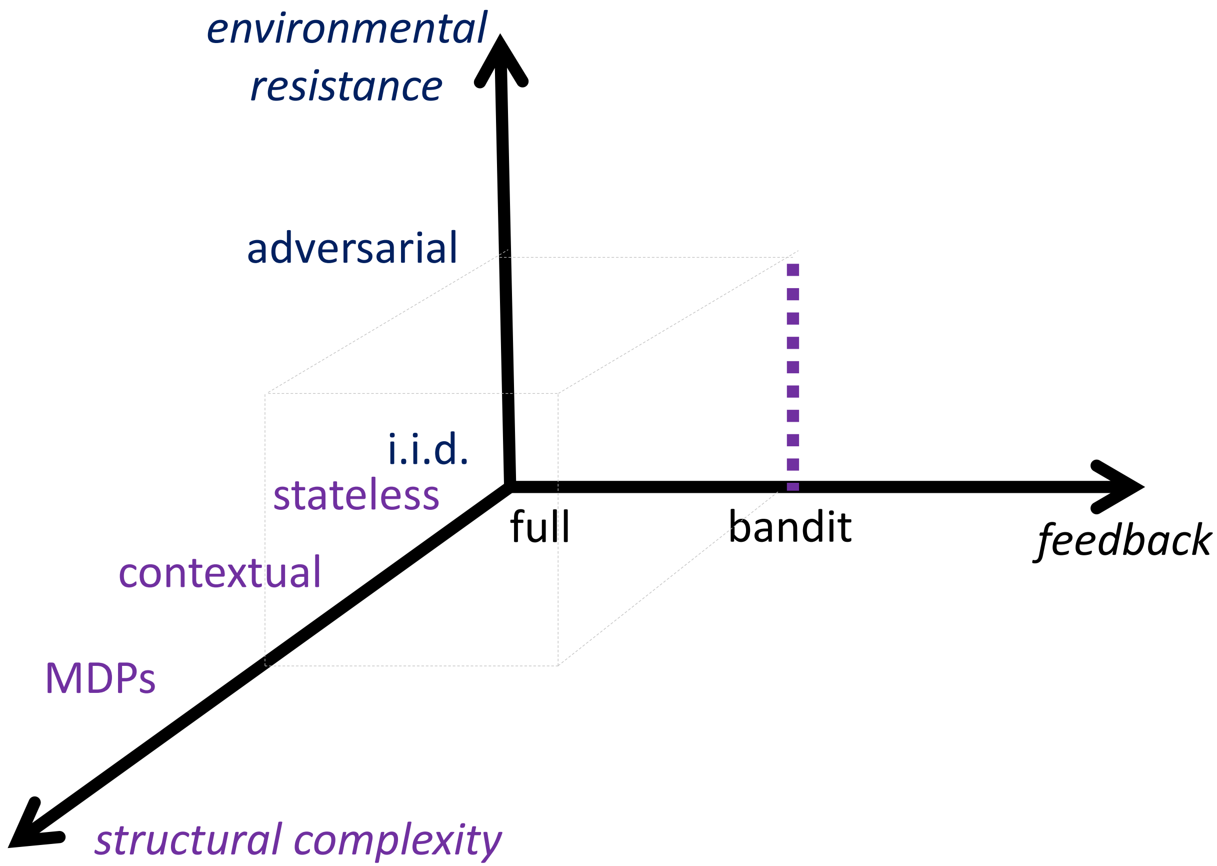
[Thune & Seldin, NIPS, 2018]



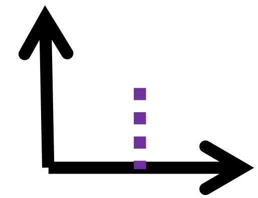
- Adversarial
  - $R_T = O\left(\varepsilon\sqrt{KT\ln K}\right)$
- Stochastic
  - $R_T = O\left(\frac{K\varepsilon^2}{\ln K}\sum_{i:\Delta_i>0}\frac{1}{\Delta_i}\right)$
- Knowledge of i.i.d./adversarial not required!
- Knowledge of ε not required!
- Simultaneous adaptation to two types of easiness!

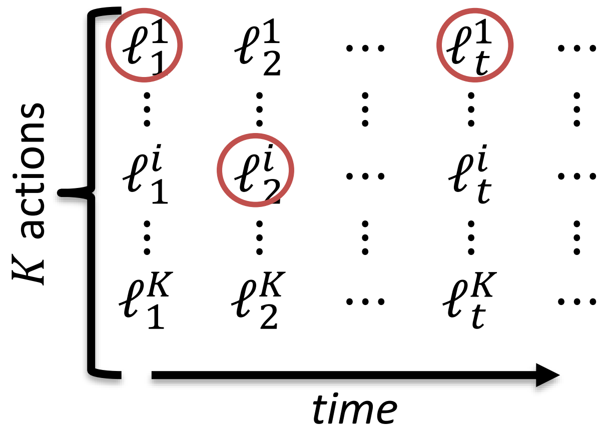
An optimal algorithm for i.i.d. and adversarial bandits

[Zimmert & Seldin, AISTATS, 2019]



I.I.D. and adversarial bandits
Problem Setting



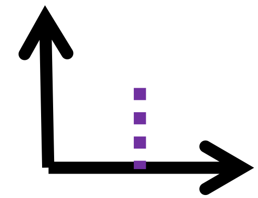


- Adversarial – as before
- Stochastically Constrained Adversary [Wei & Luo, 2018]
  - i^* – best action
  - $\mathbb{E}\left[\ell_t^i - \ell_t^{i^*}\right] = \Delta_i \geq 0$
- I.I.D.: special case when
  - $\mathbb{E}\left[\ell_t^i\right] = \mu_i$

- Evaluation measure: pseudo-regret (as before)
  - $R_T = \mathbb{E}\left[\sum_{t=1}^T \ell_t^{A_t}\right] - \min_{i \in \{1,\dots,K\}} \mathbb{E}\left[\sum_{t=1}^T \ell_t^i\right]$

α-Tsallis-INF

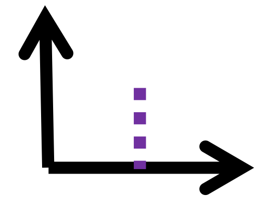
[Zimmert & Seldin, AISTATS, 2019]



- Tsallis entropy
  - $H_\alpha(x) = \frac{1}{1-\alpha}\left(1 - \sum_i x_i^\alpha\right)$
- INF – Implicitly Normalized Forecaster
  - [Audibert & Bubeck, 2009]

α-Tsallis-INF

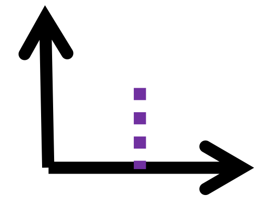
[Zimmert & Seldin, AISTATS, 2019]



- $w_t = \arg\min_{w \in \Delta^K} \langle w, \tilde{L}_{t-1} \rangle + \sum_i \frac{(w^i)^\alpha}{\alpha \eta_{t,i}}$
- Sample $A_t \sim w_t$
- Update $\tilde{L}_t^i = \tilde{L}_{t-1}^i + \frac{\ell_t^i \mathbb{I}(A_t = i)}{w_t^i}$

- $\alpha = 1$ corresponds to entropic regularization
  - EXP3 algorithm
- $\alpha = 0$ corresponds to log-barrier

$\frac{1}{2}$-Tsallis-INF: Regret Guarantees



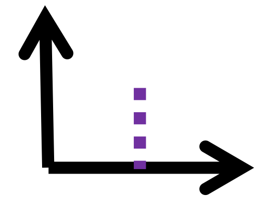
[Zimmert & Seldin, AISTATS, 2019]

- $\eta_t = \sqrt{1/t}$
- Adversarial
  - $R_T \leq 4\sqrt{KT} + 1$
- Stochastically Constrained Adversary
  - $R_T \leq \sum_{i \neq i^*} \frac{4 \ln T + 20}{\Delta_i} + 4\sqrt{K}$
  - i.i.d. is a special case

- Both results match the lower bounds (up to constants)!
- No knowledge of i.i.d./adversarial is required!

Tsallis-INF in relation to prior work

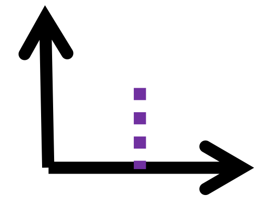
[Zimmert & Seldin, AISTATS, 2019]



	Regime	$\frac{Upper\ Bound}{Lower\ Bound}$
BROAD [Wei&Luo,2018] Log-barrier + doubling $(\alpha = 0)$	i.i.d.	$O(K^2)$
	adversarial	$O\left(\sqrt{\ln T}\right)$
$\boldsymbol{\alpha = \frac{1}{2}}$	**i.i.d. & adversarial**	$\boldsymbol{O(1)}$
EXP3++ [Seldin&Lugosi,2017] Entropic regularization + mix in extra exploration $(\alpha = 1)$	i.i.d.	$O(\ln T)$
	adversarial	$O\left(\sqrt{\ln K}\right)$

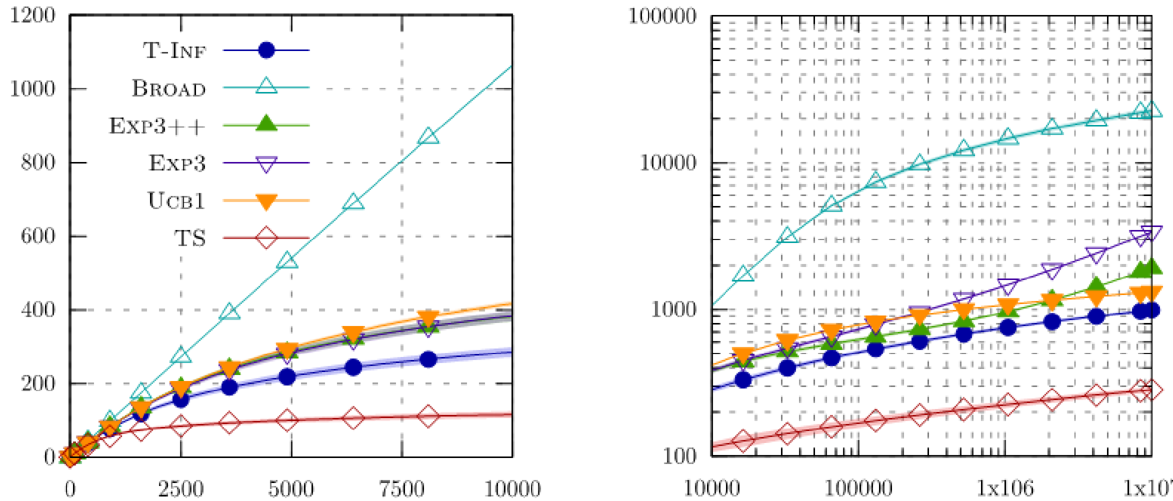
$\frac{1}{2}$-Tsallis-INF: Experiments

[Zimmert & Seldin, AISTATS, 2019]



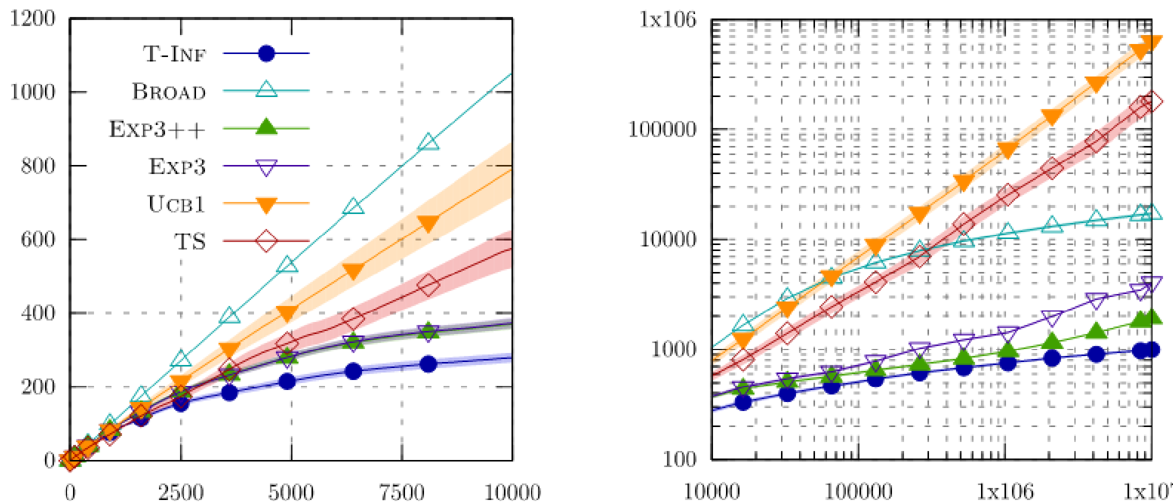
i.i.d.

a)

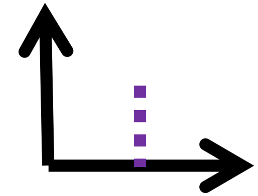


Stochastically Constrained Adversary

b)



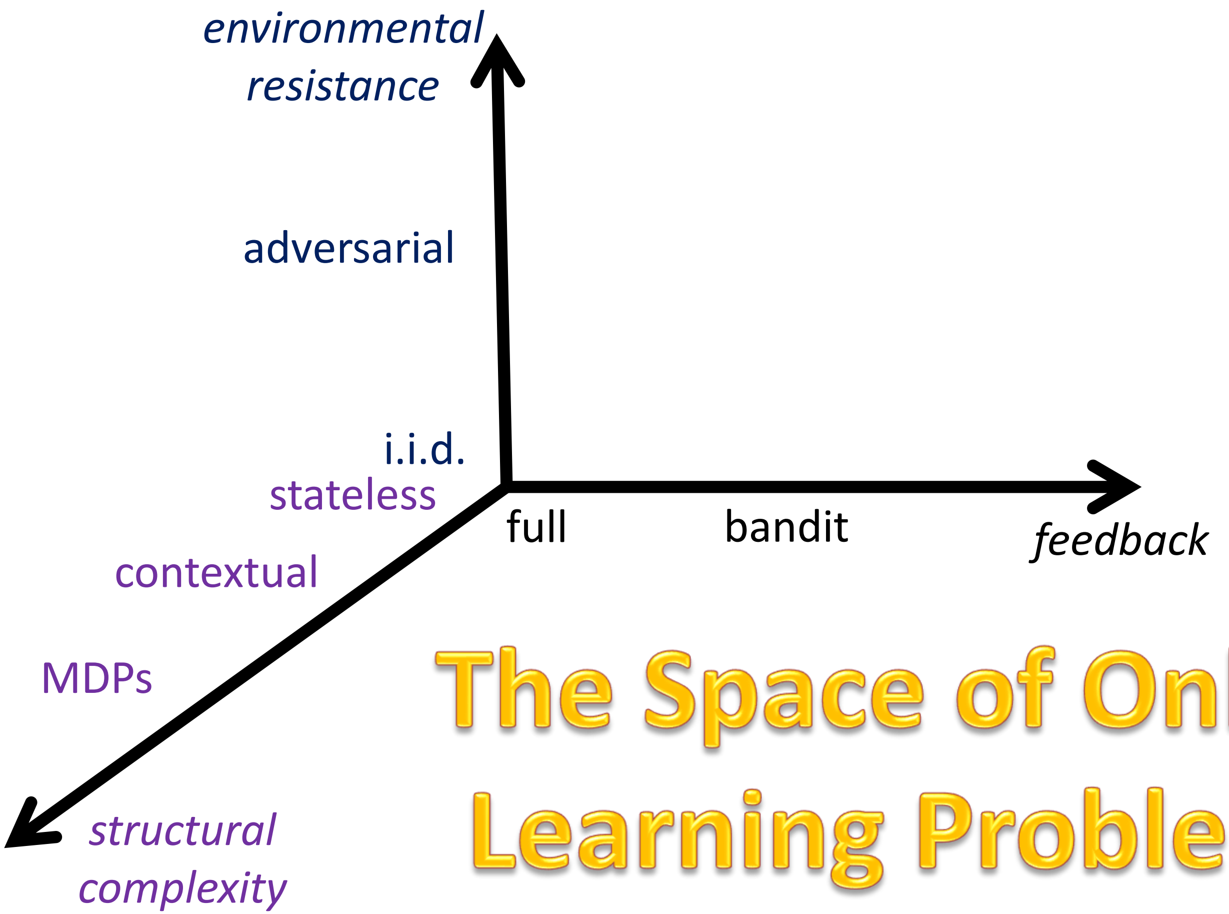
$\frac{1}{2}$-Tsallis-INF: Additional Results



[Zimmert & Seldin, AISTATS, 2019]

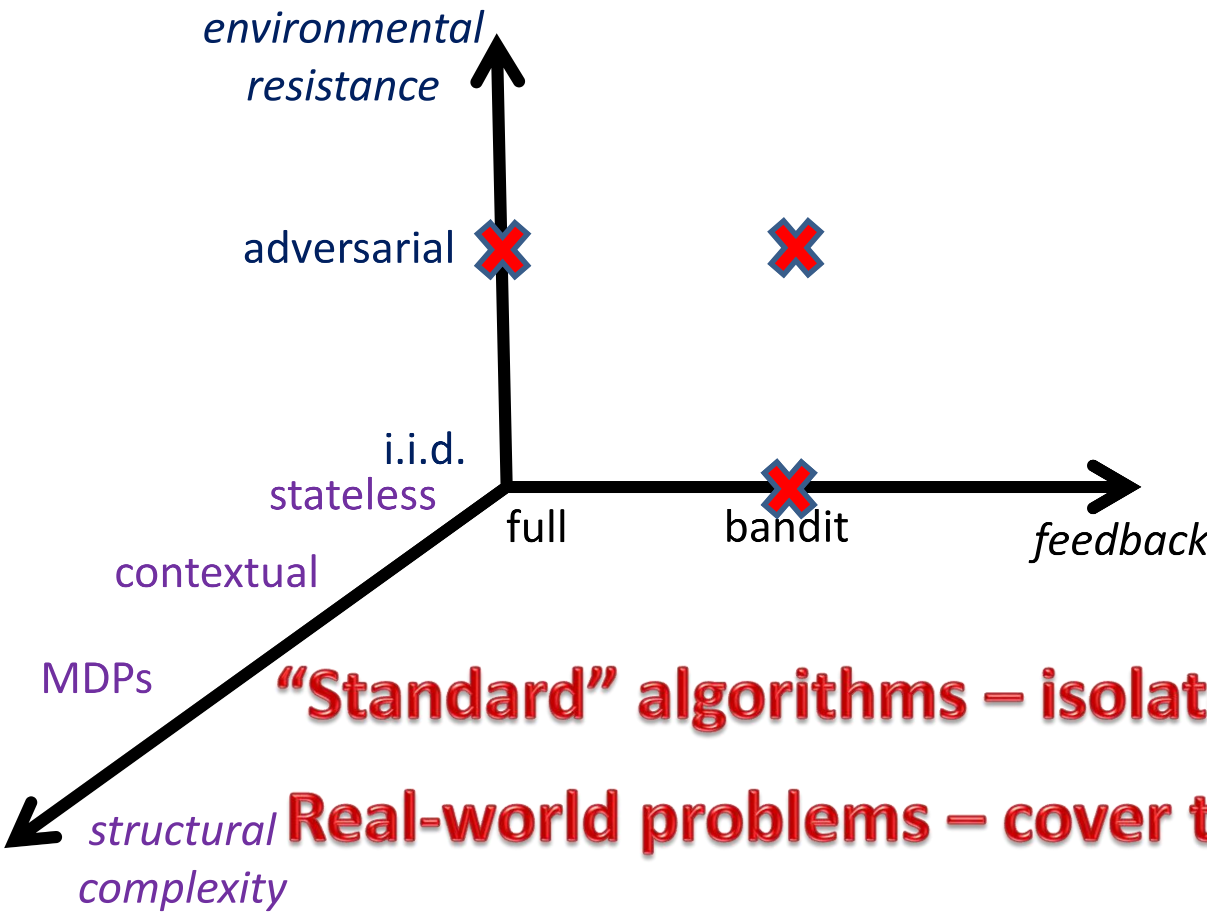
- Optimality in the moderately contaminated stochastic regime
- I.I.D. and adversarial optimality in utility-based dueling bandits

Summary



The Space of Online Learning Problems

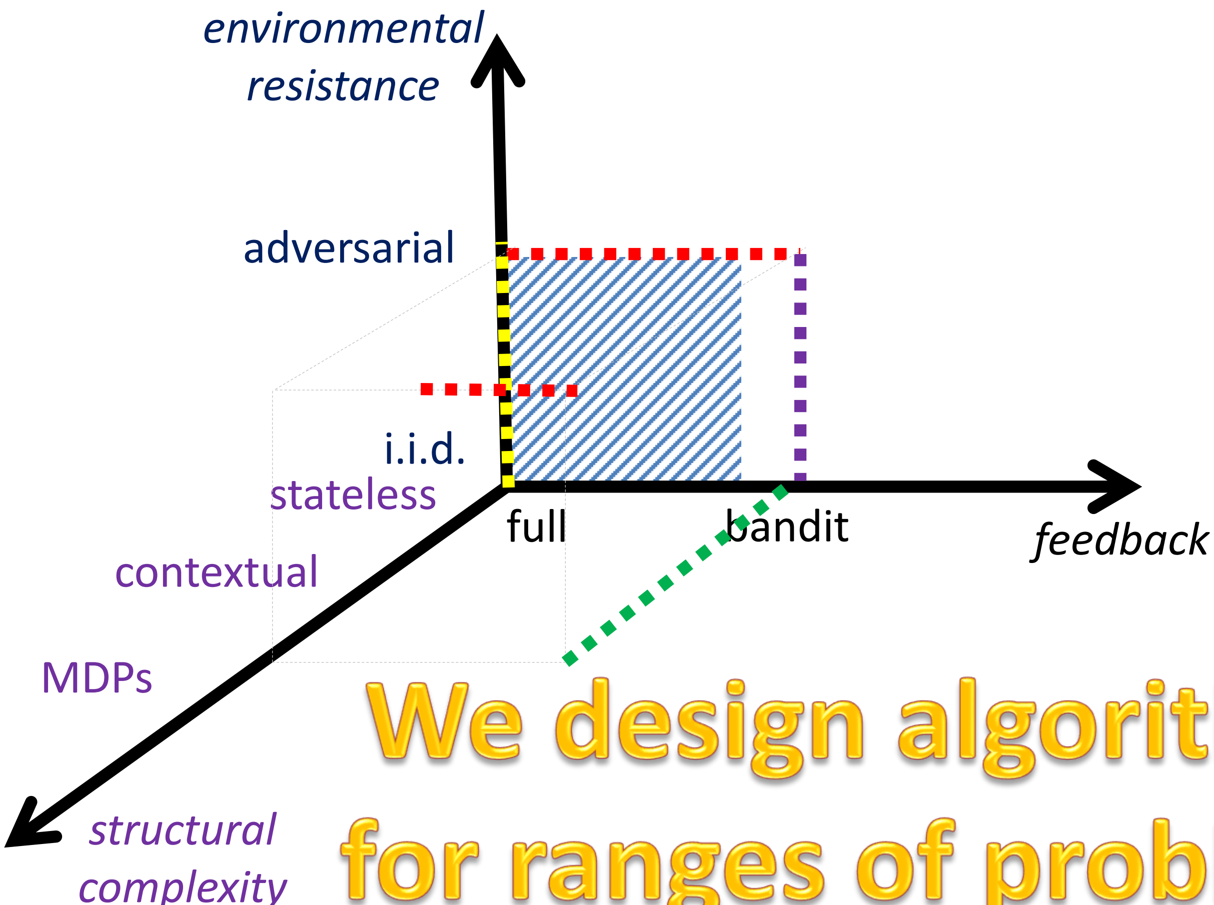
Summary



"Standard" algorithms – isolated points

Real-world problems – cover the space

Summary

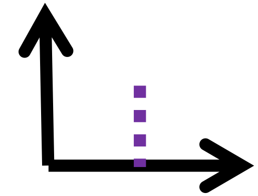


We design algorithms for ranges of problems

If you want to learn more

- Check my homepage
  - https://sites.google.com/site/yevgenyseldin/
  - (or Google me)
  - Papers, tutorials, lecture notes, etc…
- Join our Advanced Topics in Machine Learning course
  - Co-taught by Christian Igel
  - September-October 2019
  - WARNING: Math-heavy course

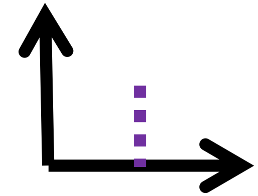
α-Tsallis-INF: Some Considerations



[Zimmert & Seldin, AISTATS, 2019]

- Exploration rate $\left(\eta_t^i\left(\tilde{L}_t^i - \tilde{L}_t^{i^*}\right)\right)^{-\frac{1}{1-\alpha}}$
- For i.i.d. regret of $\frac{\ln T}{\Delta_i}$ the exploration rate must be $\frac{1}{\Delta_i^2 t}$
- $\mathbb{E}\left[\tilde{L}_t^i - \tilde{L}_t^{i^*}\right] = \Delta_i t$
- $\left(\eta_t^i \Delta_i t\right)^{-\frac{1}{1-\alpha}} = \frac{1}{\Delta_i^2 t} \quad \Rightarrow \quad \eta_t^i = \Delta_i^{1-2\alpha} t^{-\alpha}$
- $\alpha = \frac{1}{2}$ is the only value for which η_t^i requires no tuning by (unknown) Δ_i

α-Tsallis-INF: Some Considerations



[Zimmert & Seldin, AISTATS, 2019]

- $\mathbb{E}\left[\tilde{L}_t^i - \tilde{L}_t^{i^*}\right] = \Delta_i t$
- The variance of $\tilde{L}_t^i - \tilde{L}_t^{i^*}$ is of order $\Delta_i^2 t^2$
- $\tilde{L}_t^i - \tilde{L}_t^{i^*}$ cannot be efficiently controlled by Bernstein's inequality
- Our analysis is based on a self-bounding property of the regret