Apple Tasting Revisited

An Online Binary Classification Problem with Partial Information

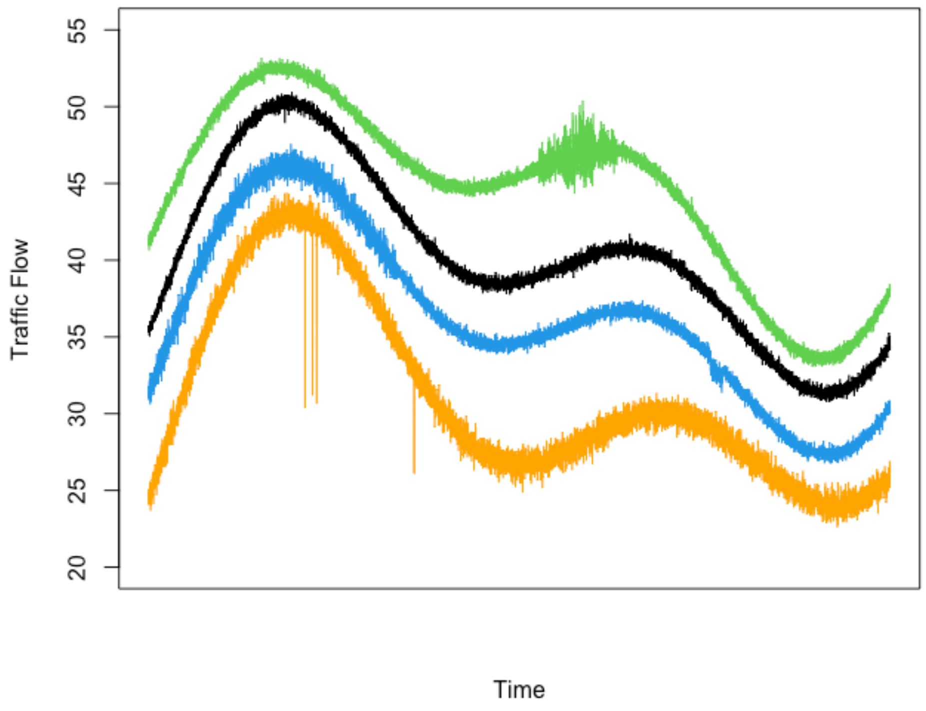
James A Grant (he/him) **-** Lancaster University

Joint work with David S Leslie

Cardiff OR and Statistics Seminar - 17th February 2022

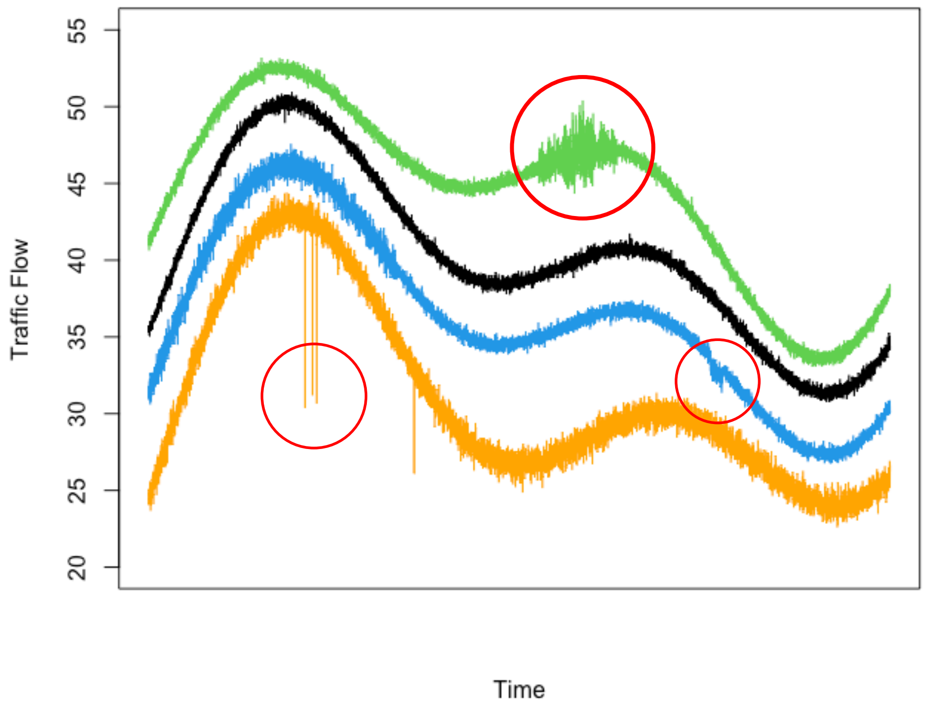
j.grant@lancaster.ac.uk - @james_a_grant

Motivation: Telecoms Network Control



Engineers monitor traffic data for outages, faults, etc. and reroute traffic or schedule maintenance accordingly.

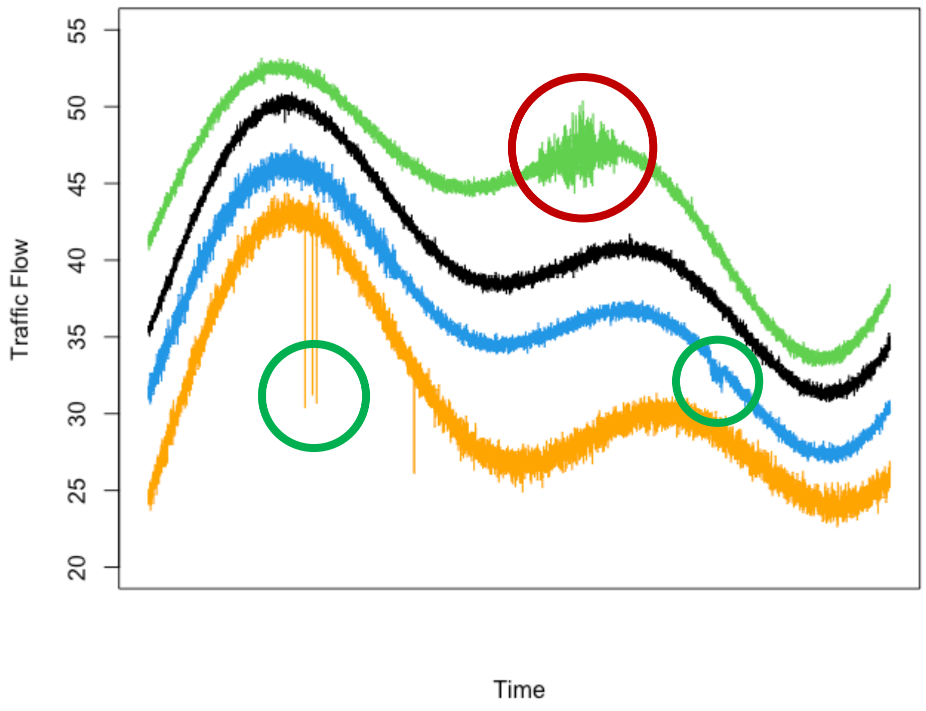
Motivation: Telecoms Network Control



Engineers monitor traffic data for outages, faults, etc. and reroute traffic or schedule maintenance accordingly.

Often aided by statistical techniques.

Motivation: Telecoms Network Control

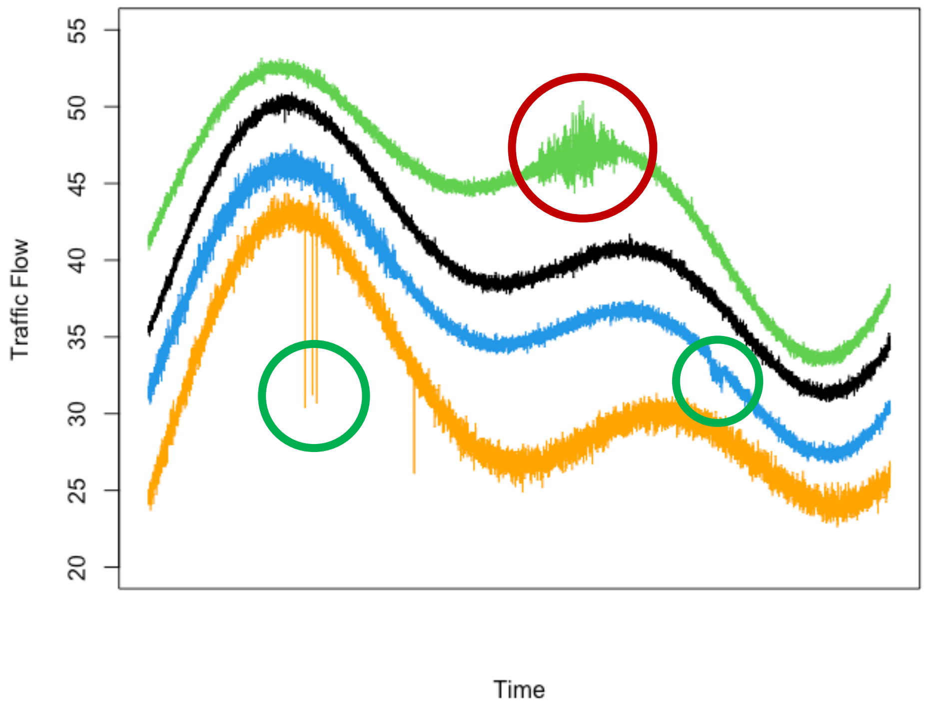


Engineers monitor traffic data for outages, faults, etc. and reroute traffic or schedule maintenance accordingly.

Often aided by statistical techniques.

Some artifacts are of genuine concern, some are innocuous.

Motivation: Telecoms Network Control

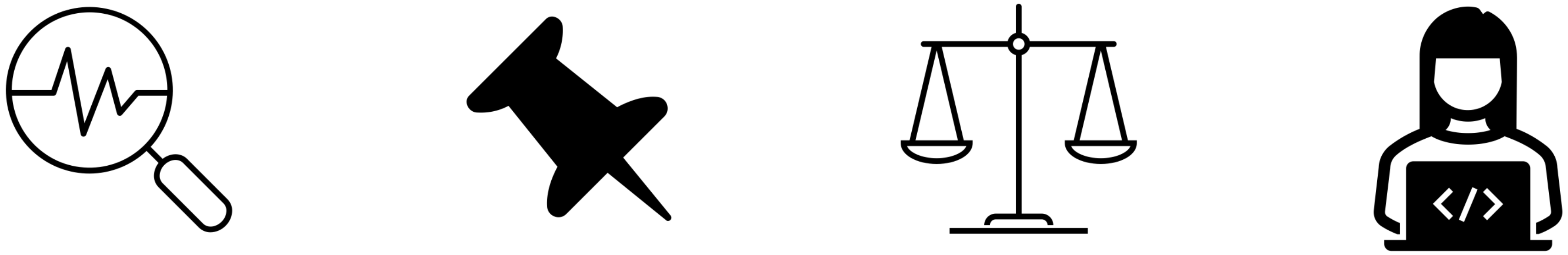


Automating this process is **hard**

- Combining different knowledge
- Domain expertise
- Actions taken are complex
- Unseen examples and changing ‘normal’ behaviour

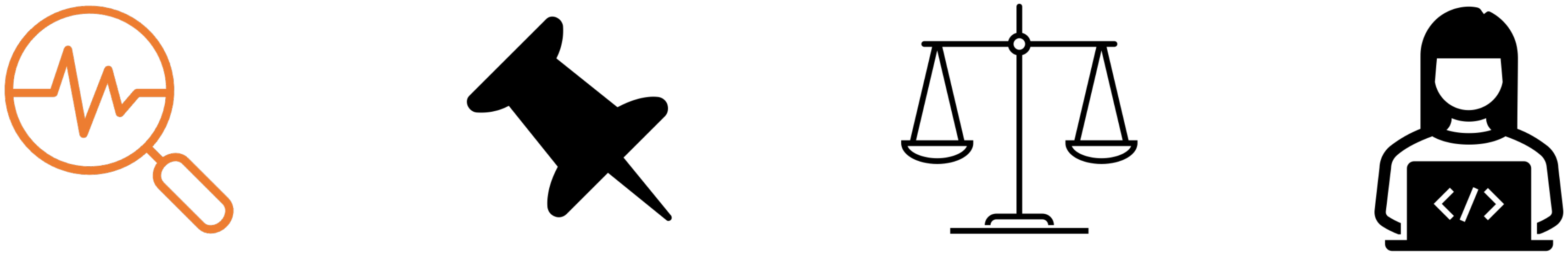
A semi-autonomous approach

We instead consider not trying to **make** decisions (per se), but flagging **when** a non-trivial decision needs to be made.



A semi-autonomous approach

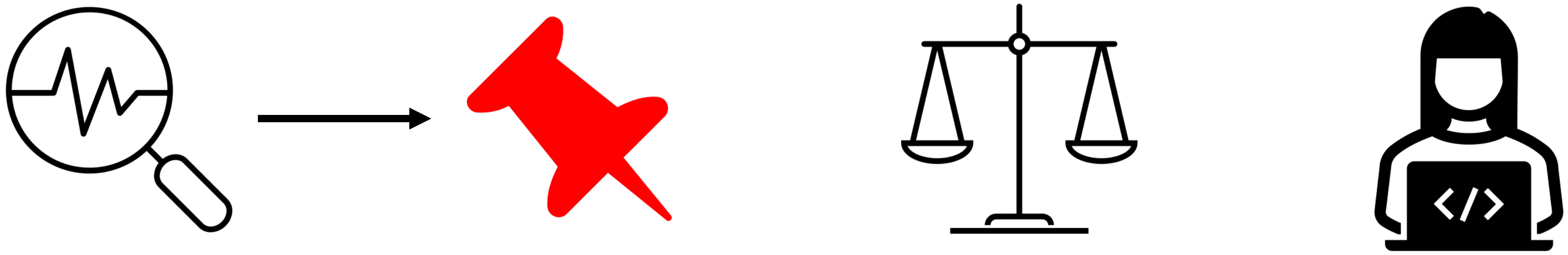
We instead consider not trying to **make** decisions (per se), but flagging **when** a non-trivial decision needs to be made.



Monitor the data

A semi-autonomous approach

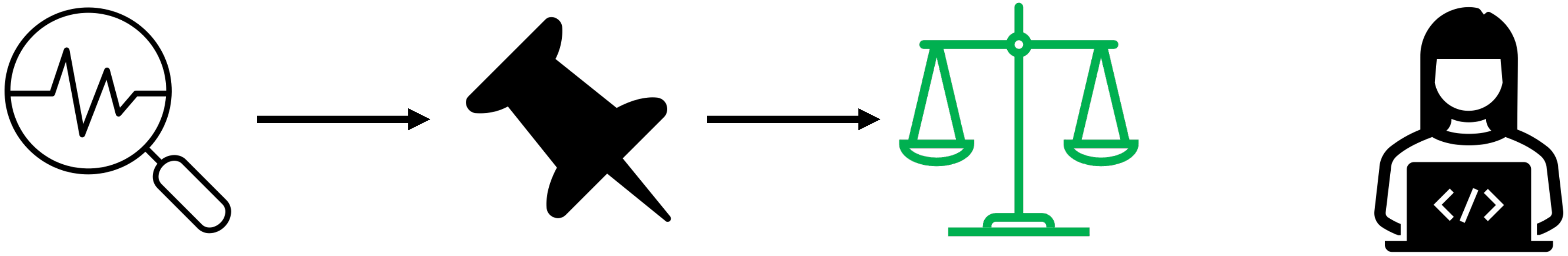
We instead consider not trying to **make** decisions (per se), but flagging **when** a non-trivial decision needs to be made.



Pin-point interesting regions

A semi-autonomous approach

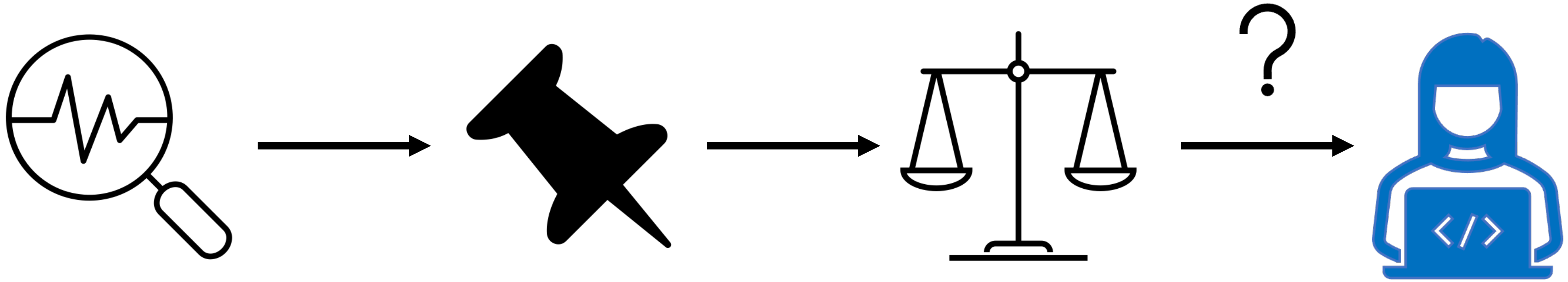
We instead consider not trying to **make** decisions (per se), but flagging **when** a non-trivial decision needs to be made.



Weigh up whether they are important

A semi-autonomous approach

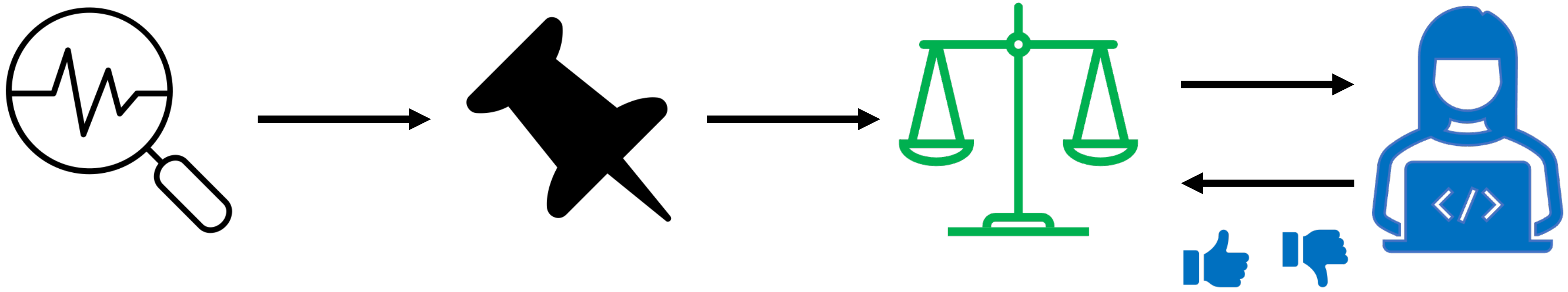
We instead consider not trying to **make** decisions (per se), but flagging **when** a non-trivial decision needs to be made.



Potentially pass to a human

A semi-autonomous approach

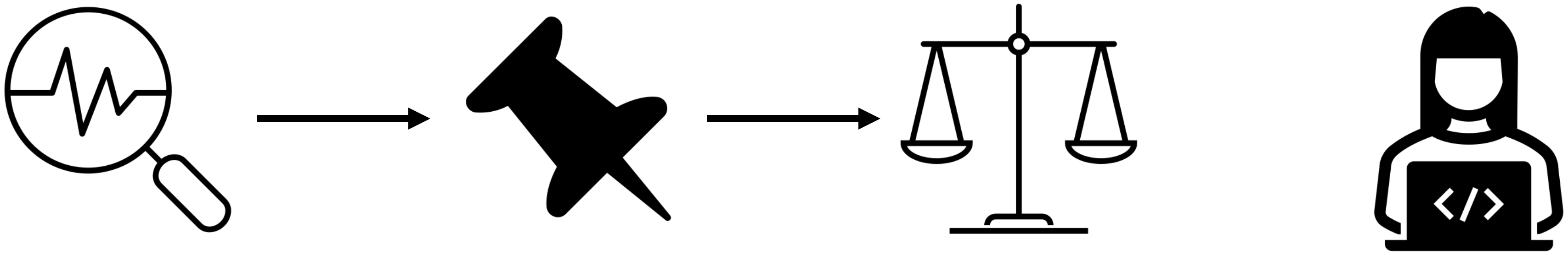
We instead consider not trying to **make** decisions (per se), but flagging **when** a non-trivial decision needs to be made.



If so, get feedback

A semi-autonomous approach

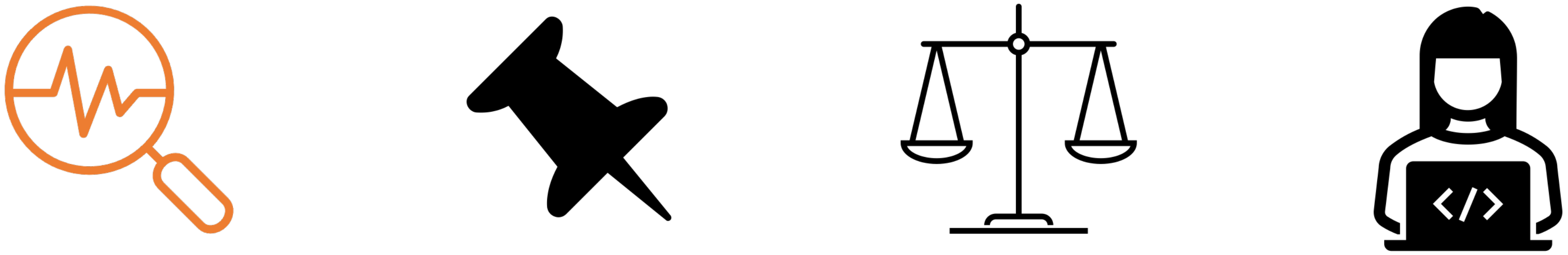
We instead consider not trying to **make** decisions (per se), but flagging **when** a non-trivial decision needs to be made.



If not, no feedback

A semi-autonomous approach

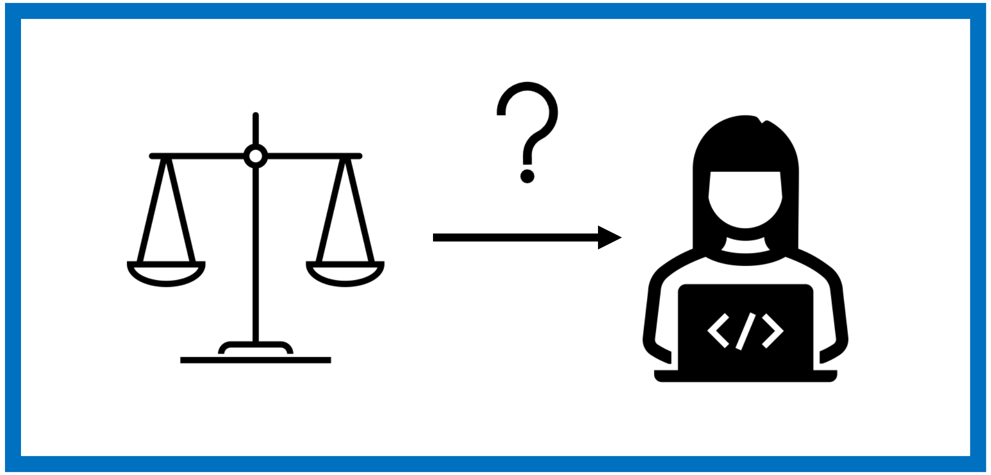
We instead consider not trying to **make** decisions (per se), but flagging **when** a non-trivial decision needs to be made.



Return to the monitoring phase

A semi-autonomous approach

We instead consider not trying to **make** decisions (per se), but flagging **when** a non-trivial decision needs to be made.



Today's Focus – Classification with Partial Feedback

Learning to Classify

Pose the decision to flag as a binary classification task.

Each potentially interesting anomaly ($t = 1, 2, \ldots$) has

- Associated feature vector $x_t \in \mathbb{R}^d$ - size of deviation/extraneous variables/baseline deviated from/etc.
- True (initially latent) class $C_t \in \{0,1\}$ – not interesting/interesting

To some extent x_t's can predict C_t's – e.g. logistic regression-like relationship mediated by parameter $\theta \in \mathbb{R}^d$,

$$C_t \sim Bern\left(\sigma(x_t^T \theta)\right).$$

Learning to Classify

Offline Binary Classification: Have a history of $x_1, \ldots x_n$ and $C_1, \ldots, C_n$ and produce estimate $\hat{\theta}_n$. Predict any future $\hat{C}_t$ based on x_t and $\hat{\theta}_n$.

Online Binary Classification: Little or no historic data. Iteratively observe x_t, predict $\hat{C}_t$, observe **true** C_t, and update estimate $\hat{\theta}_t$.

Online Binary Classification with Partial Feedback: Same setting as online – but only observe true C_t if $\hat{C}_t = 1$.

Online Binary Classification with Partial Feedback, or '**Apple Tasting**'.

Apple Tasting

- Learning to identify good and bad apples *(Helmbold et al. 1992, 2000)*.
- **Aim**: let all good apples through, remove all bad apples.
- Class only revealed by taste – which destroys the apple:
    - Desirable for bad apples. Wasteful for good apples.

Apple Tasting

- Learning to identify good and bad apples *(Helmbold et al. 1992, 2000)*.
- **Aim**: let all good apples through, remove all bad apples.
- Class only revealed by taste – which destroys the apple:
  - Desirable for bad apples. Wasteful for good apples.

- Challenge is that to maximise accuracy, some good apples must be removed for sake of learning – **but which ones and how many**?

Balancing Exploration and Exploitation

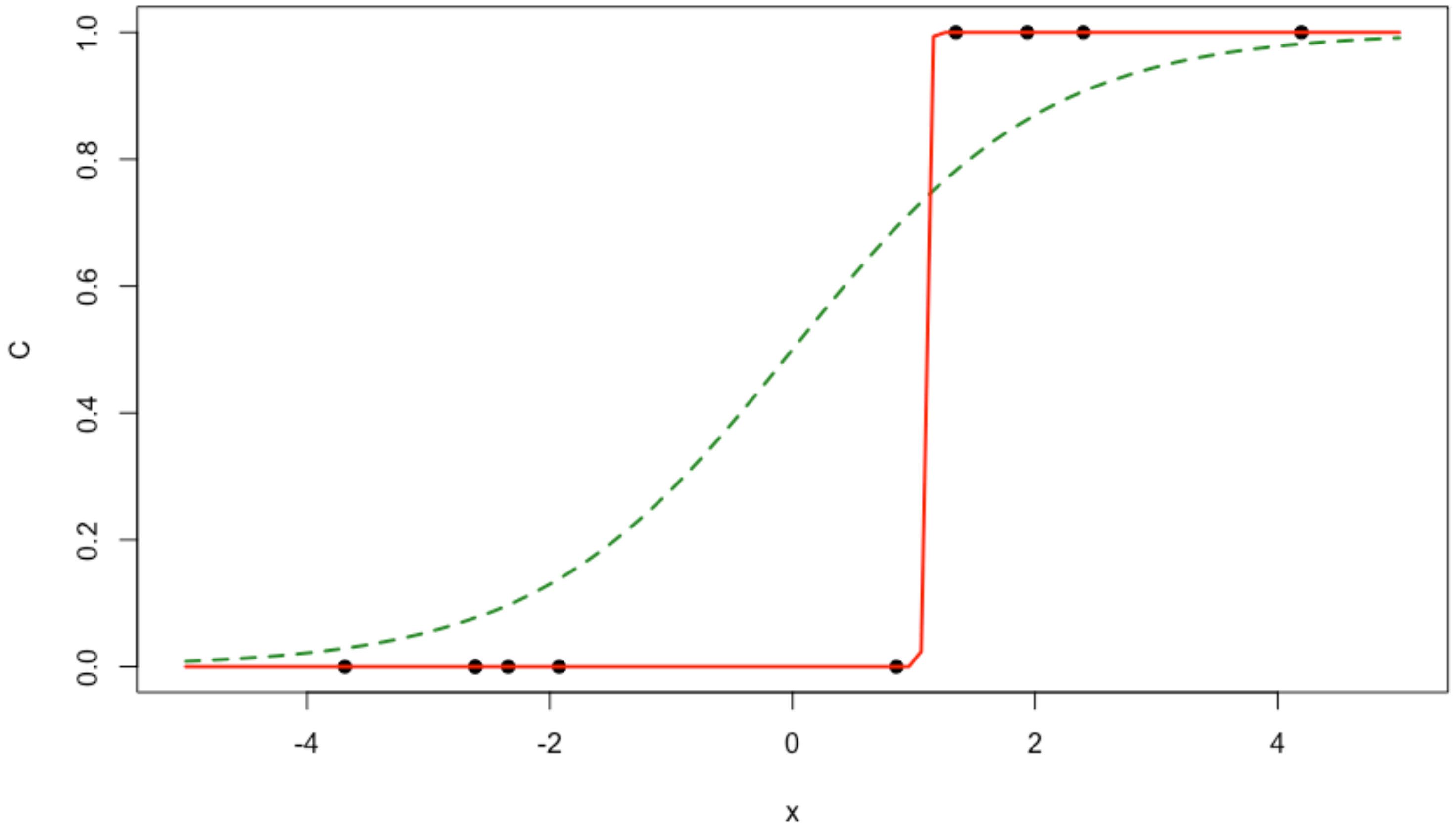
- Repeatedly face the following question:
  - Given observed features x_t, and a guess of the class $P(C_t = 1)$ (based on a $\hat{\theta}_t$) do we choose treat as a good or bad apple?

Balancing Exploration and Exploitation

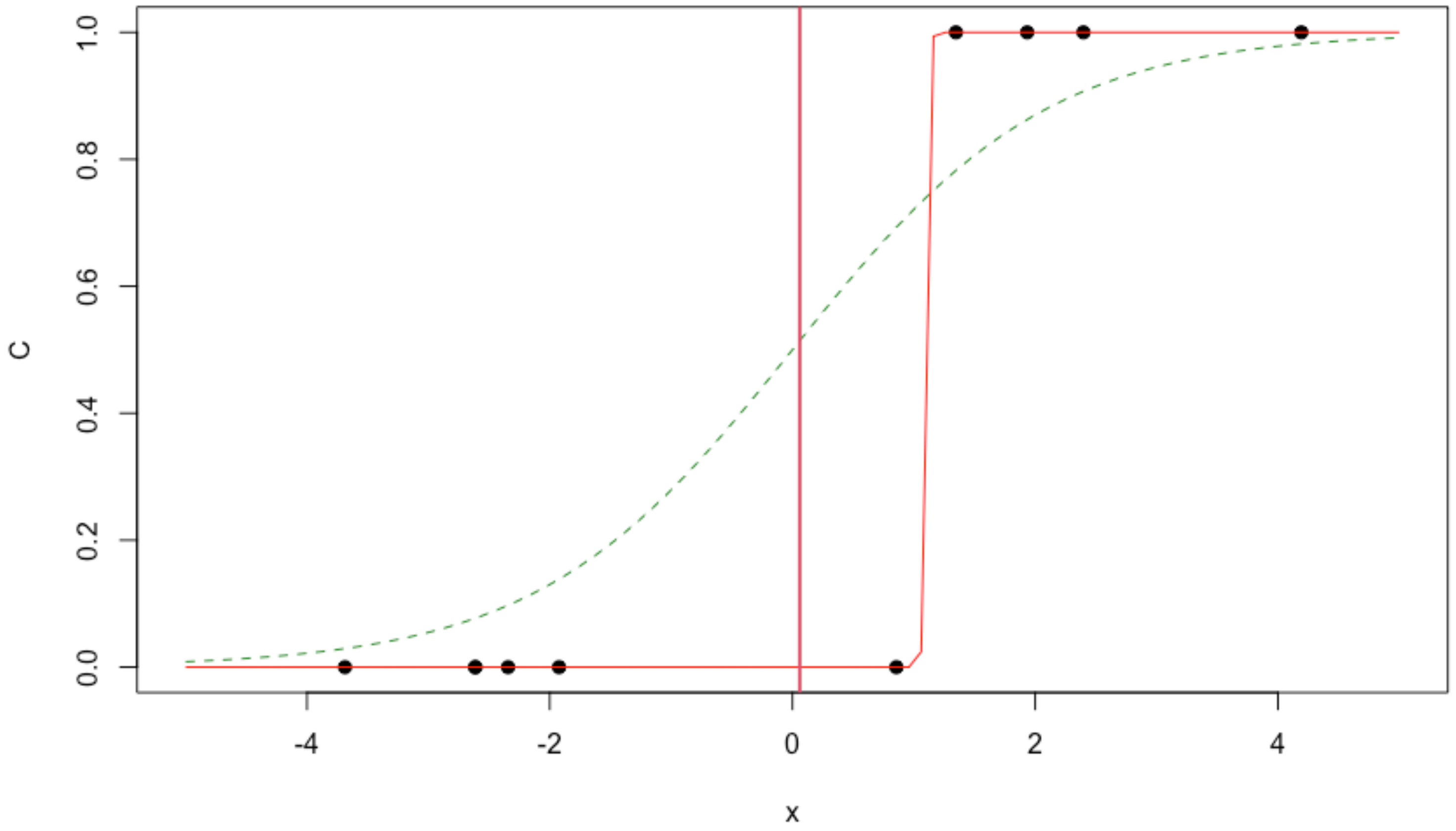
- Repeatedly face the following question:
  - Given observed features x_t, and a guess of the class $P(C_t = 1)$ (based on a $\hat{\theta}_t$) do we choose treat as a good or bad apple?

- Why not just use best guess all the time?
  - Could work brilliantly - if x_i sequence is sufficiently variable, if you start with good data
  - Could also fail catastrophically – initialise $\hat{\theta}$ poorly and only observe data which confirms bias.

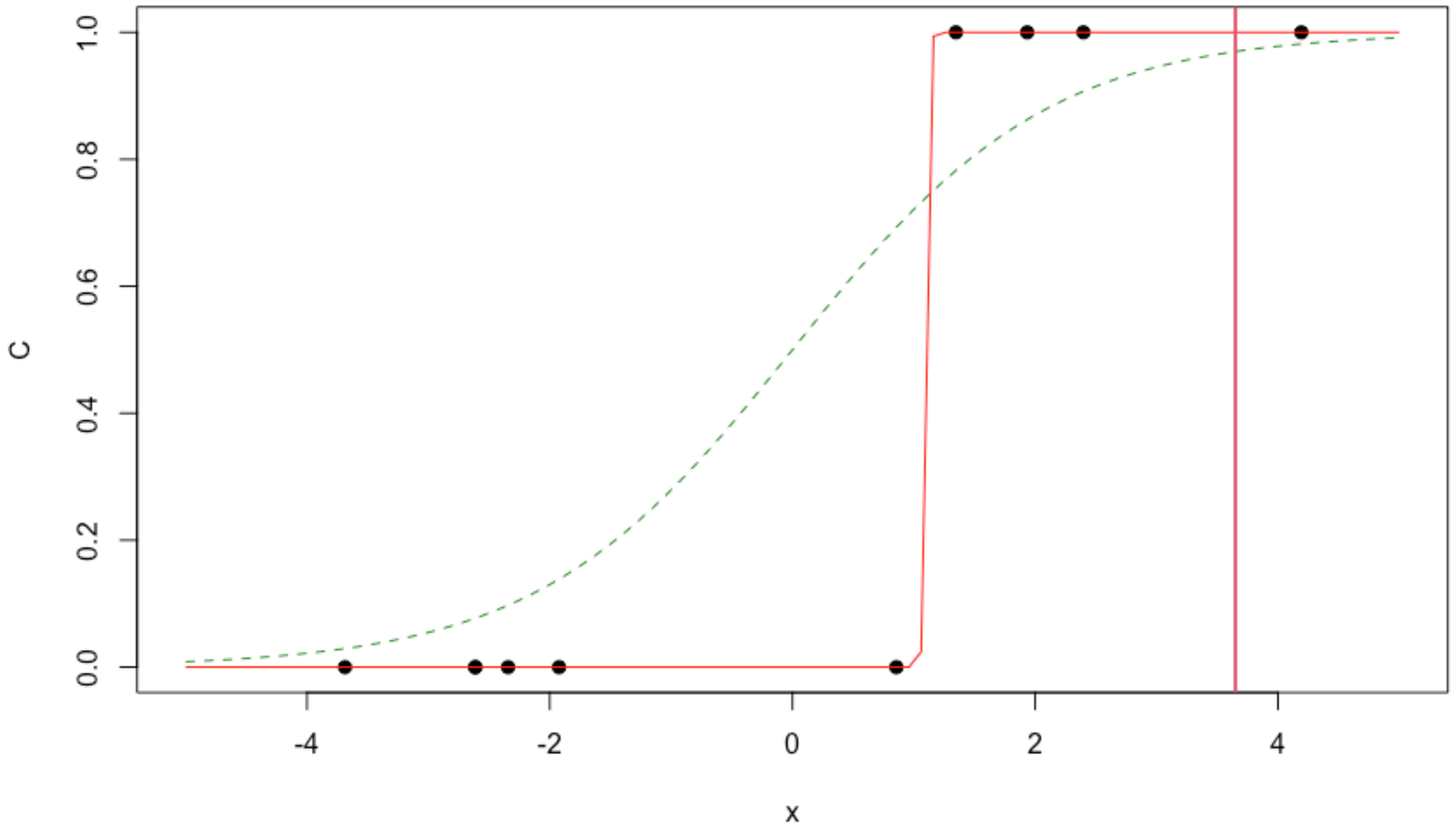
Balancing Exploration and Exploitation



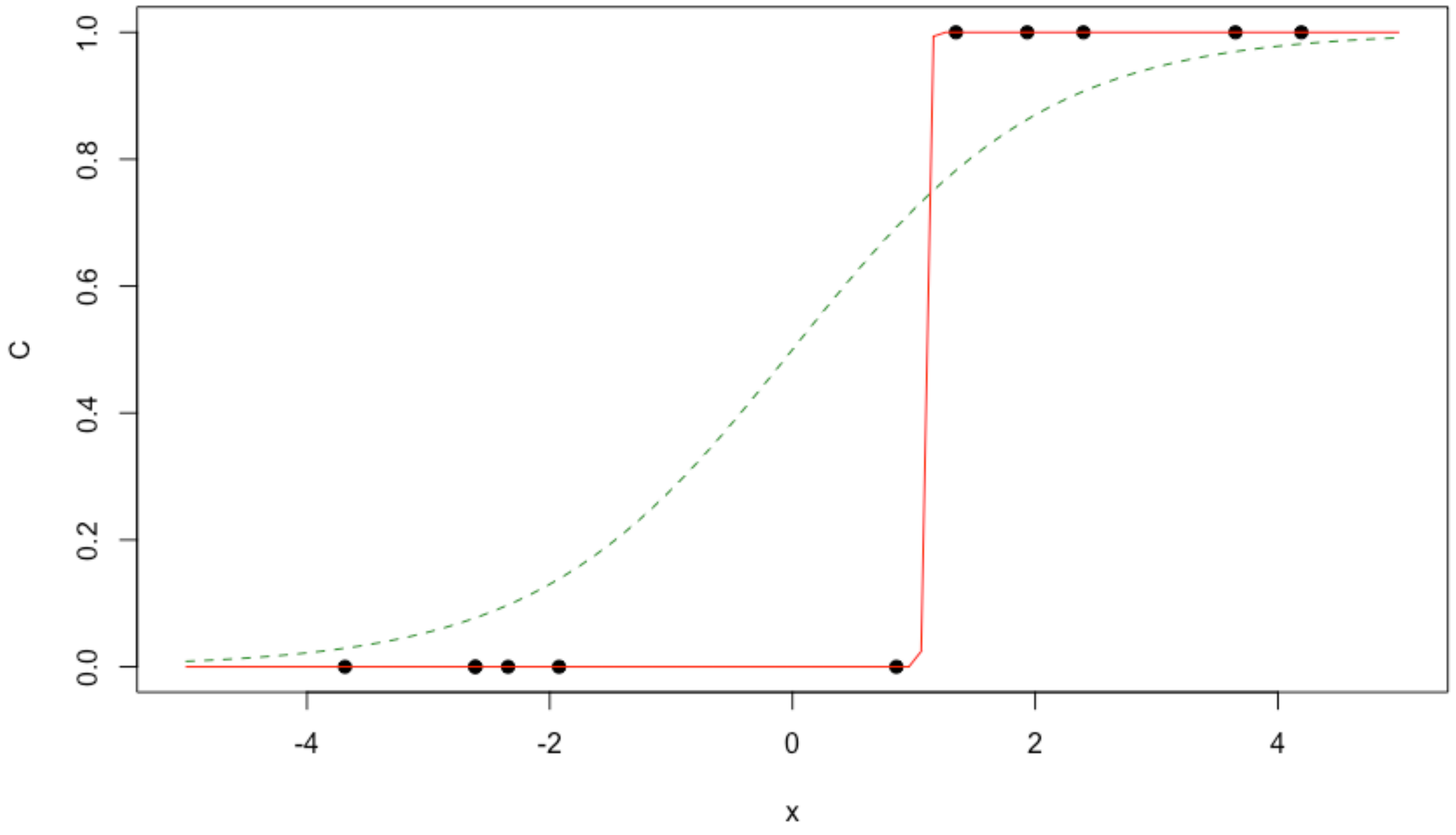
Balancing Exploration and Exploitation



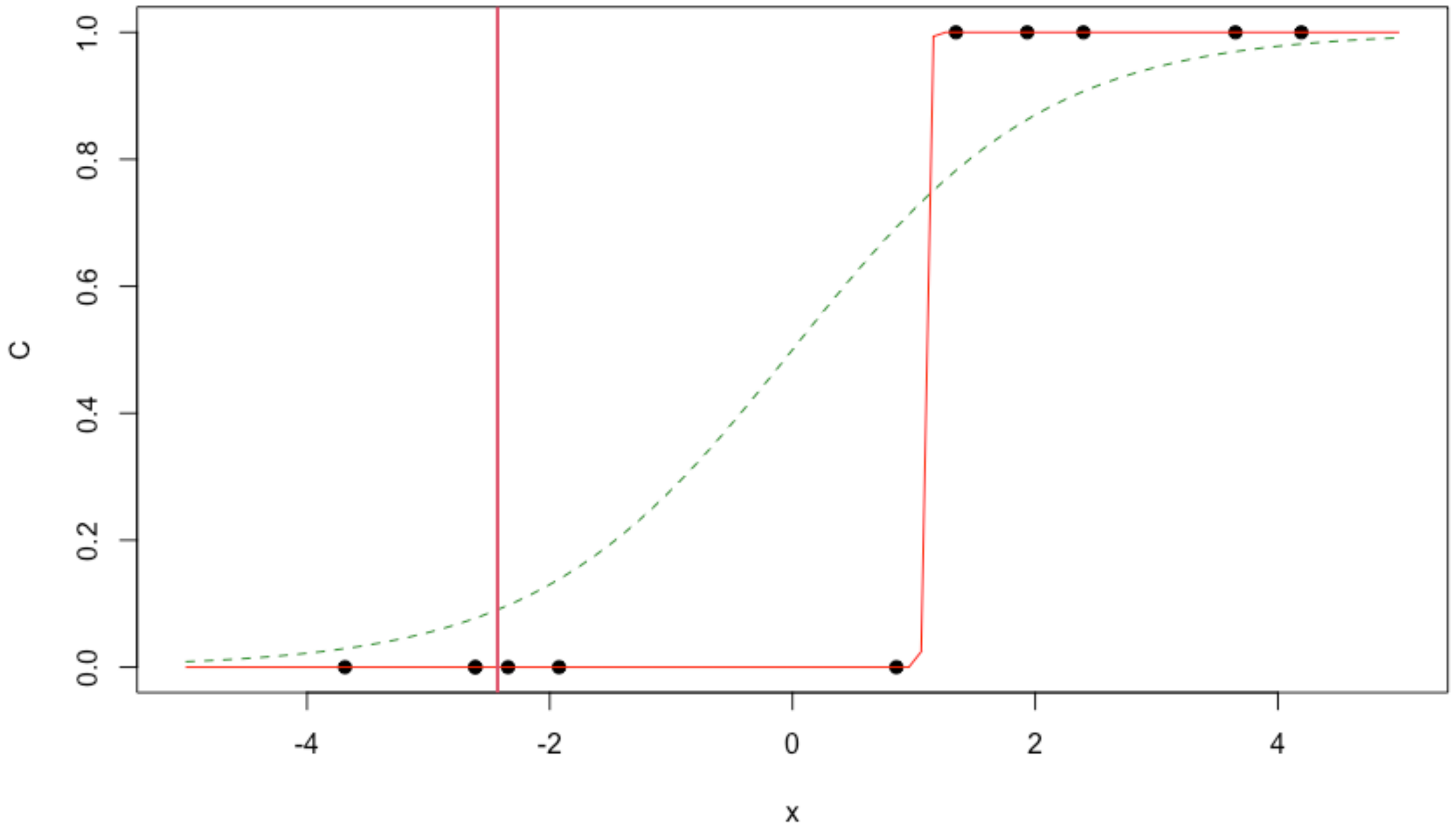
Balancing Exploration and Exploitation



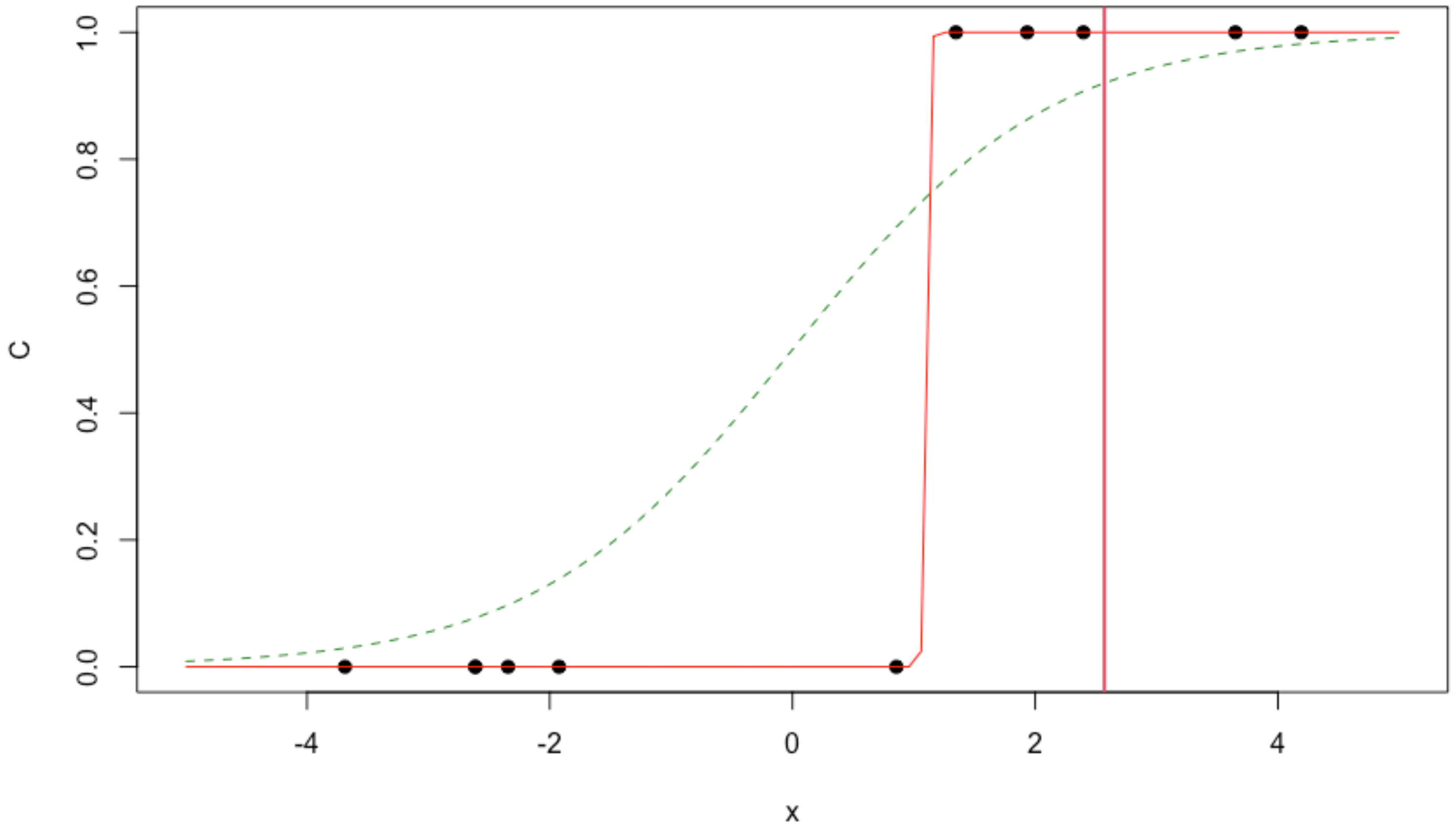
Balancing Exploration and Exploitation



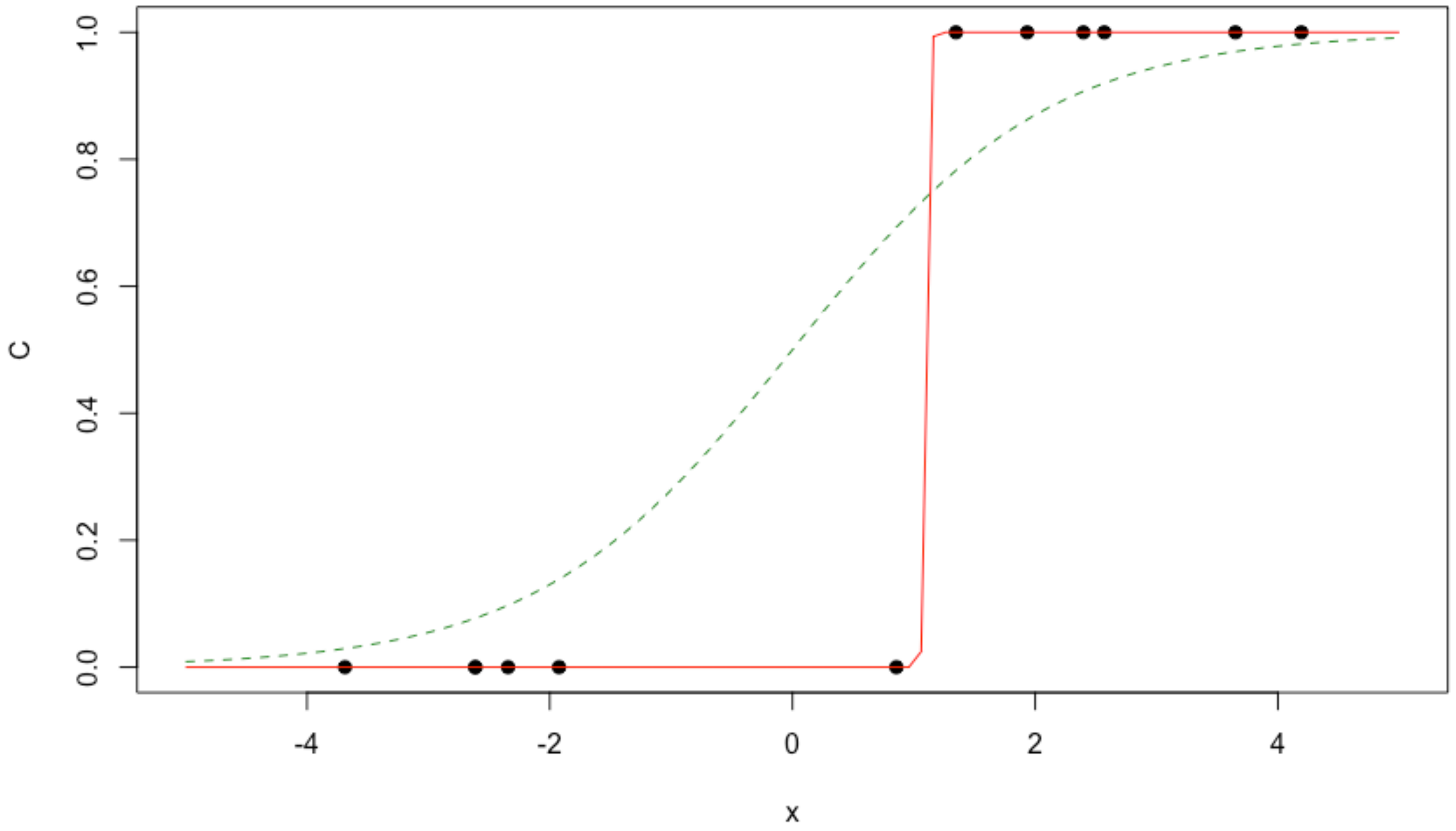
Balancing Exploration and Exploitation



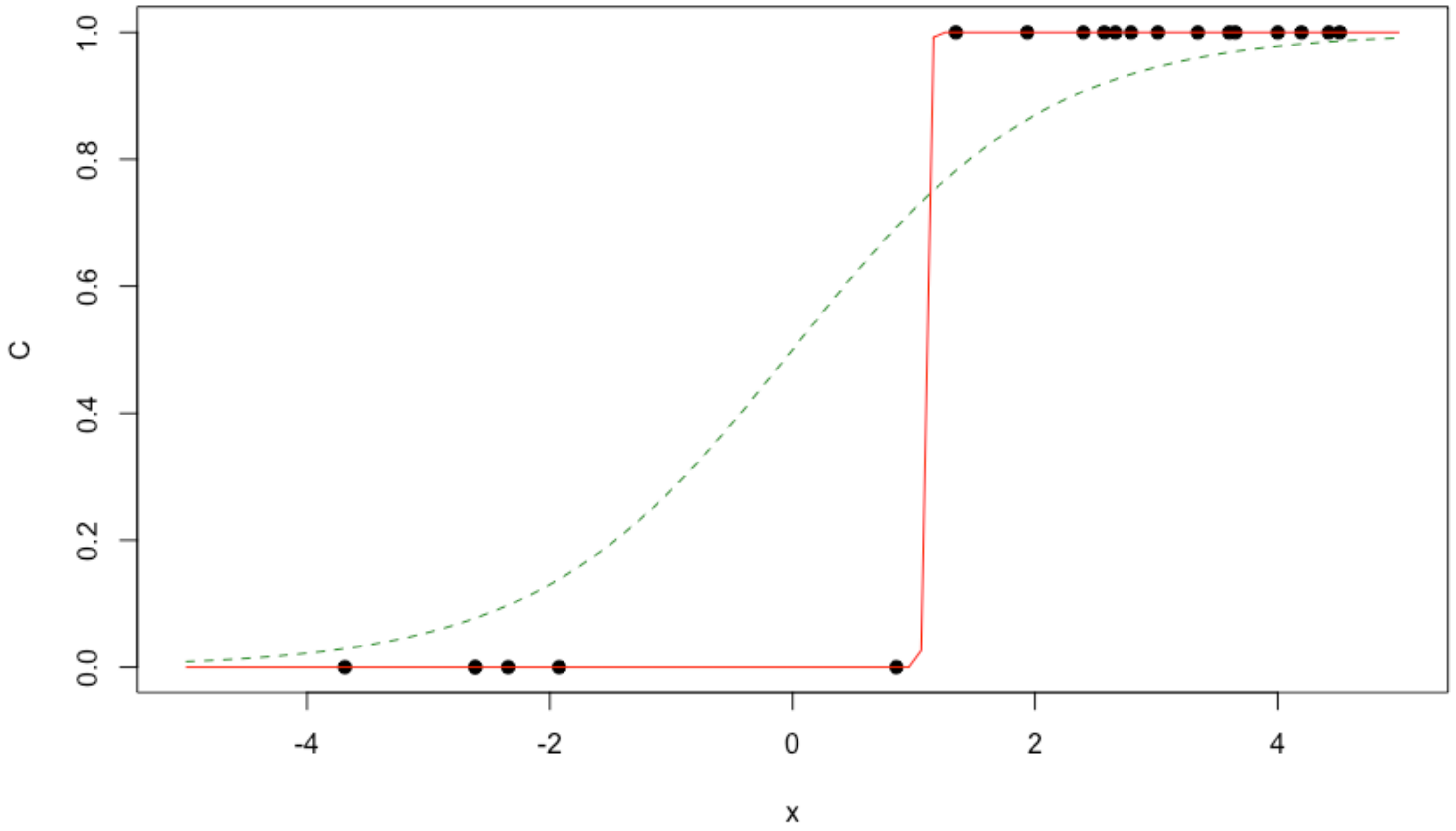
Balancing Exploration and Exploitation



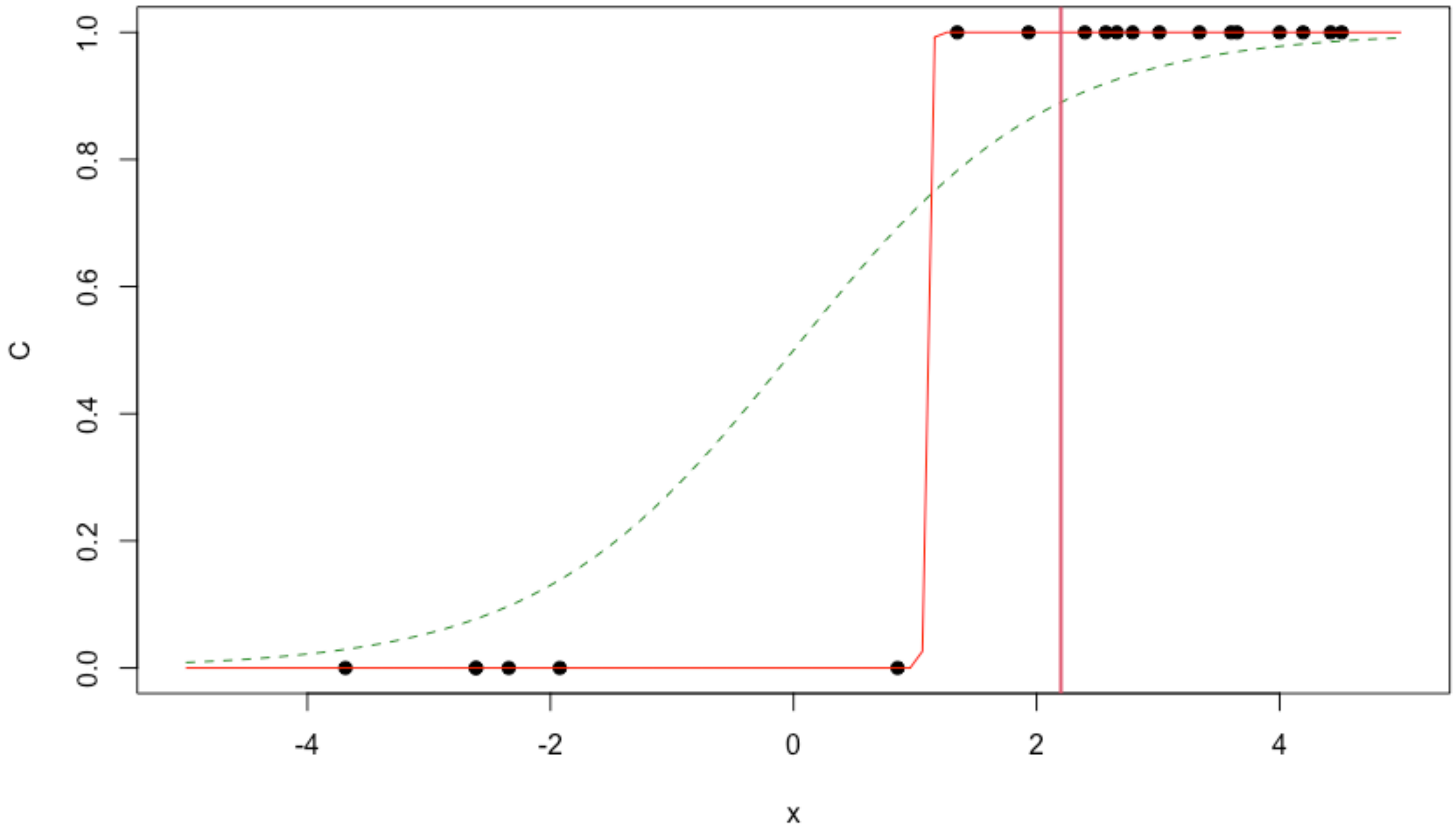
Balancing Exploration and Exploitation



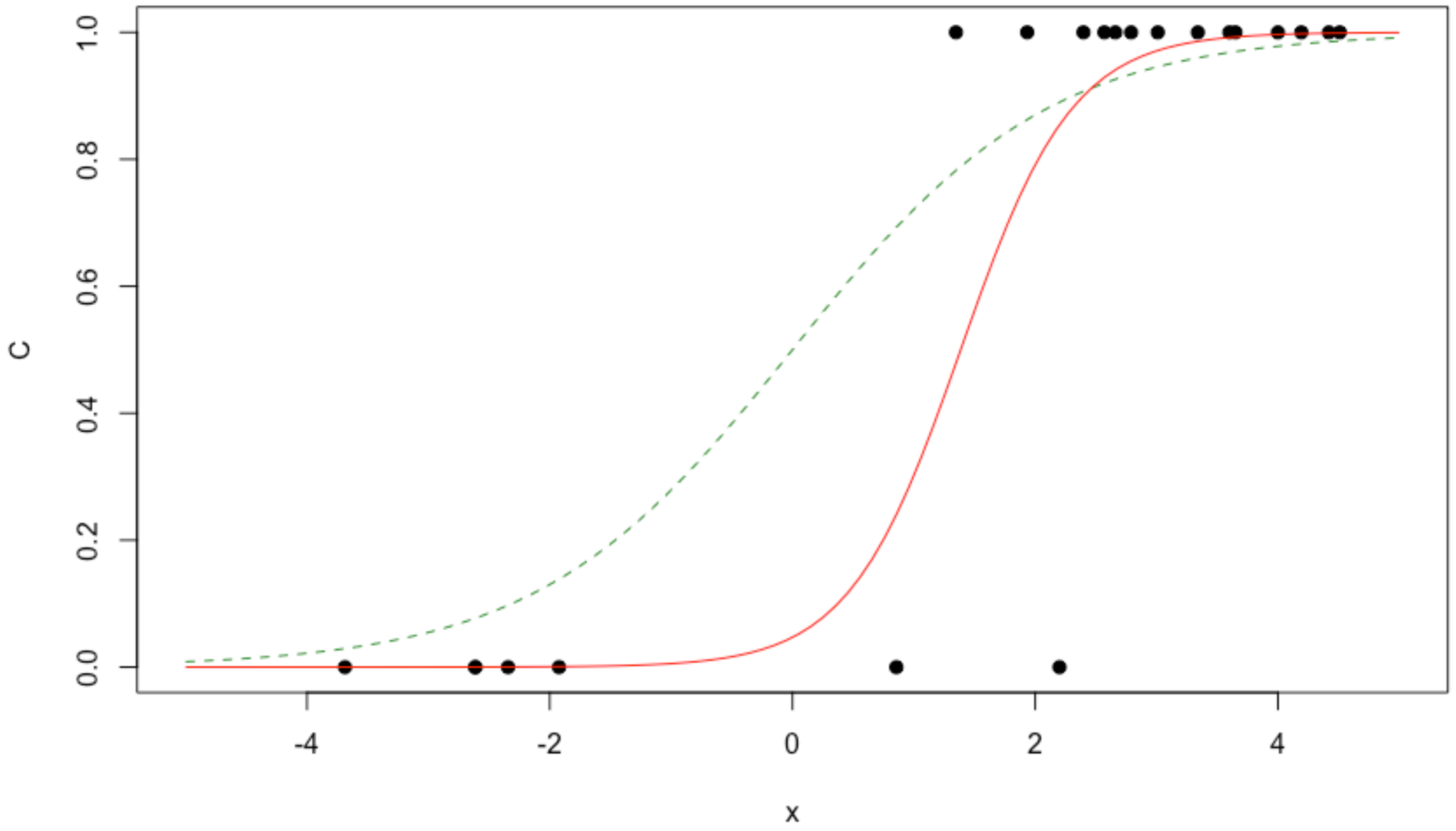
Balancing Exploration and Exploitation



Balancing Exploration and Exploitation



Balancing Exploration and Exploitation



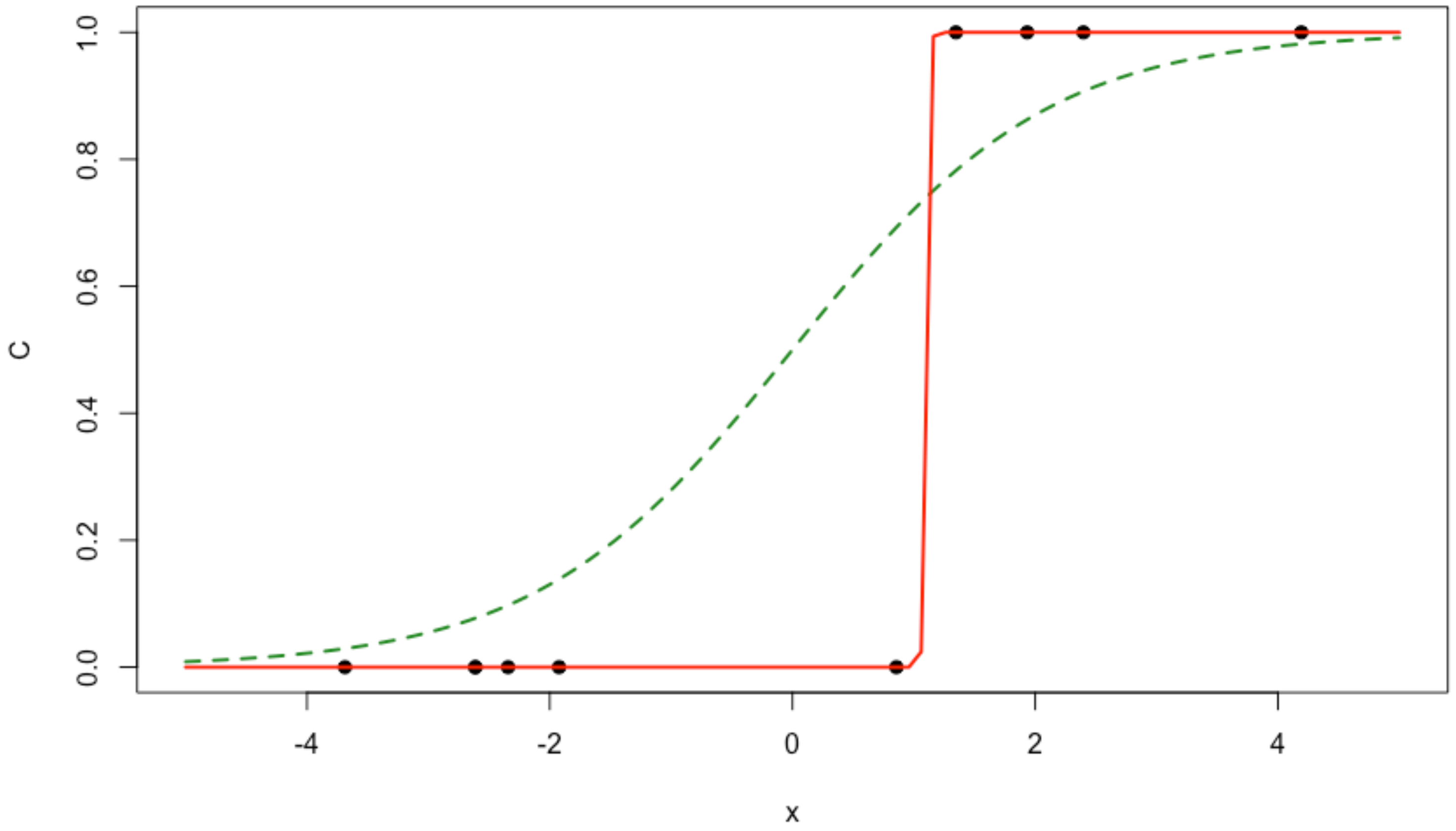
Balancing Exploration and Exploitation

- Superior methods ensure we have enough data to maintain a good estimate of $\hat{\theta}_t$.
- Two main techniques:
  - **Confidence bounds** - only treat as a good apple if we're very certain it's good (effectively shift $\hat{\theta}_t$ to the limit of some region Θ_t such that $P(\theta \in \Theta_t) > 1 - \delta$)
  - **Randomisation** – add (appropriate) noise to $\hat{\theta}_t$, so that sometimes an estimated label $\hat{C}_t$ will be flipped (encouraging exploration)
- Both converge to using $\hat{C}_t$ once $\hat{\theta}_t$ is well estimated.

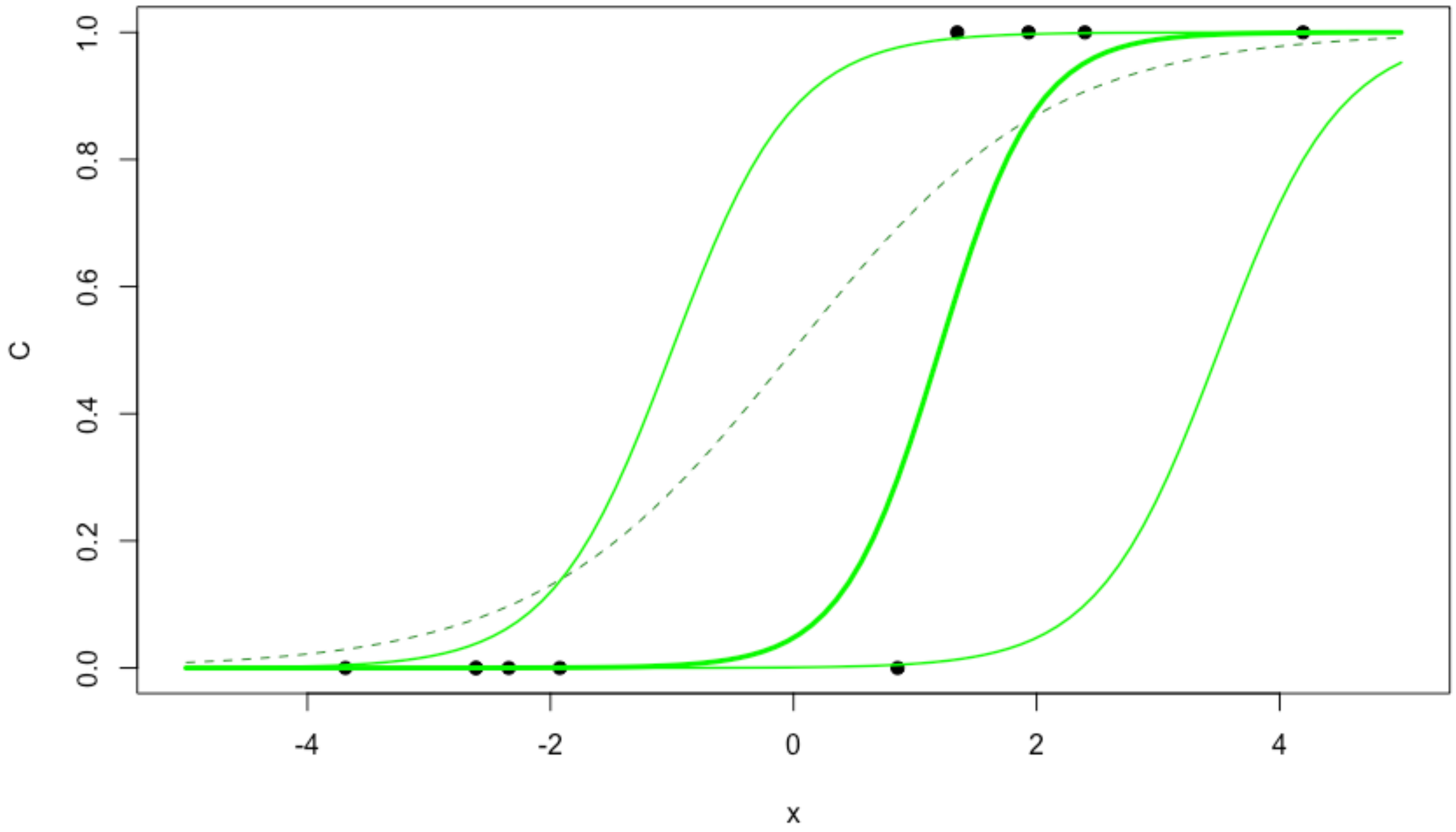
Thompson Sampling

- Initialise with a prior distribution $\pi_0(\theta)$
- At time $t = 1, 2, \ldots$
  - Draw a sample $\tilde{\theta}_t$ from the current posterior $\pi_{t-1}(\theta)$
  - Treat $\tilde{\theta}_t$ as the true parameter and estimate $\hat{C}(\tilde{\theta}_t)$ based on x_t.
  - If $\hat{C}(\tilde{\theta}_t) = 1$
    - Remove the apple/show anomaly to human
    - Observe C_t and update the belief distribution to $\pi_t(\theta)$.
  - If $\hat{C}(\tilde{\theta}_t) = 0$
    - Let apple/anomaly pass
    - Observe nothing and set $\pi_t(\theta) = \pi_{t-1}(\theta)$.

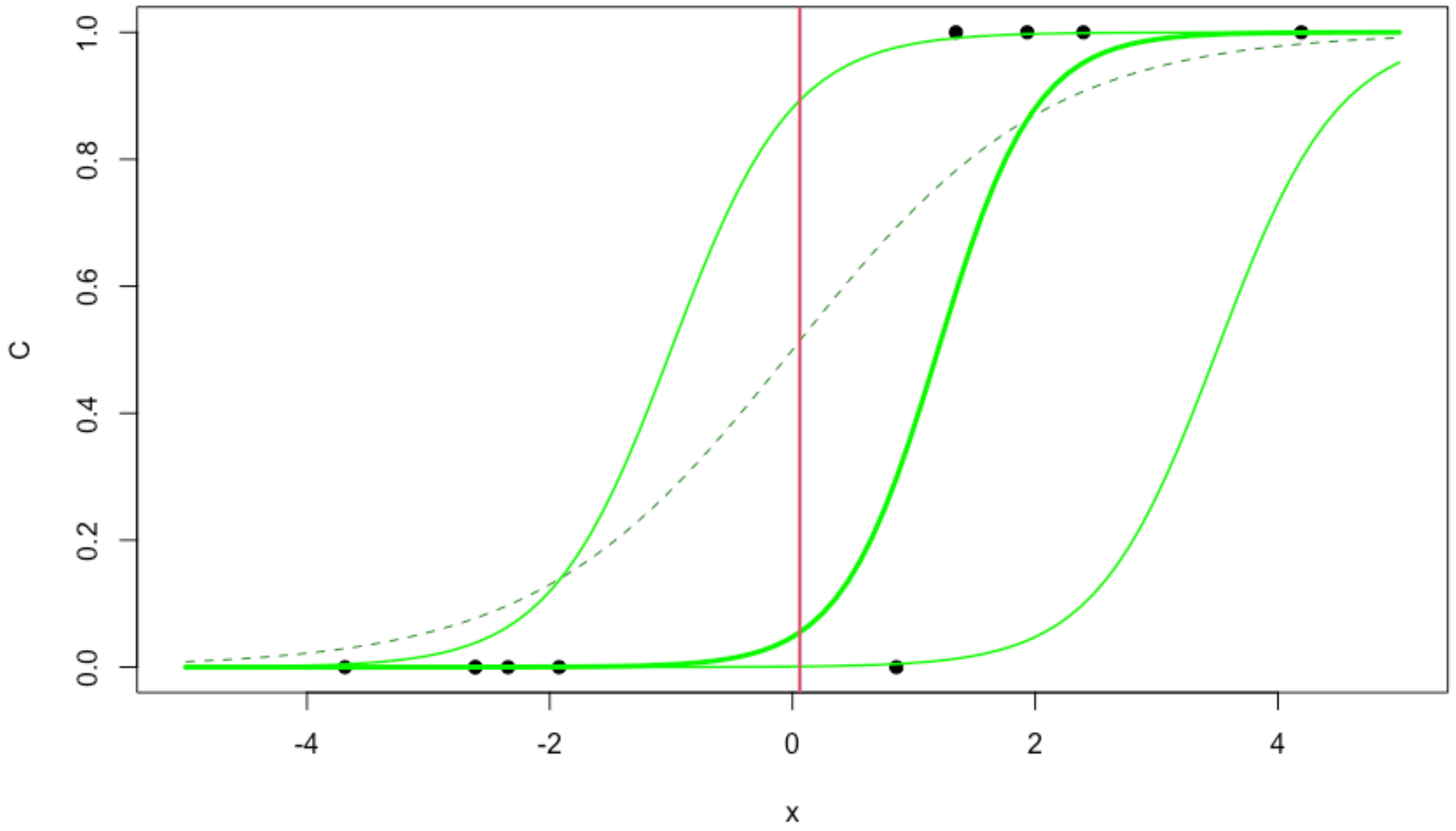
Thompson Sampling



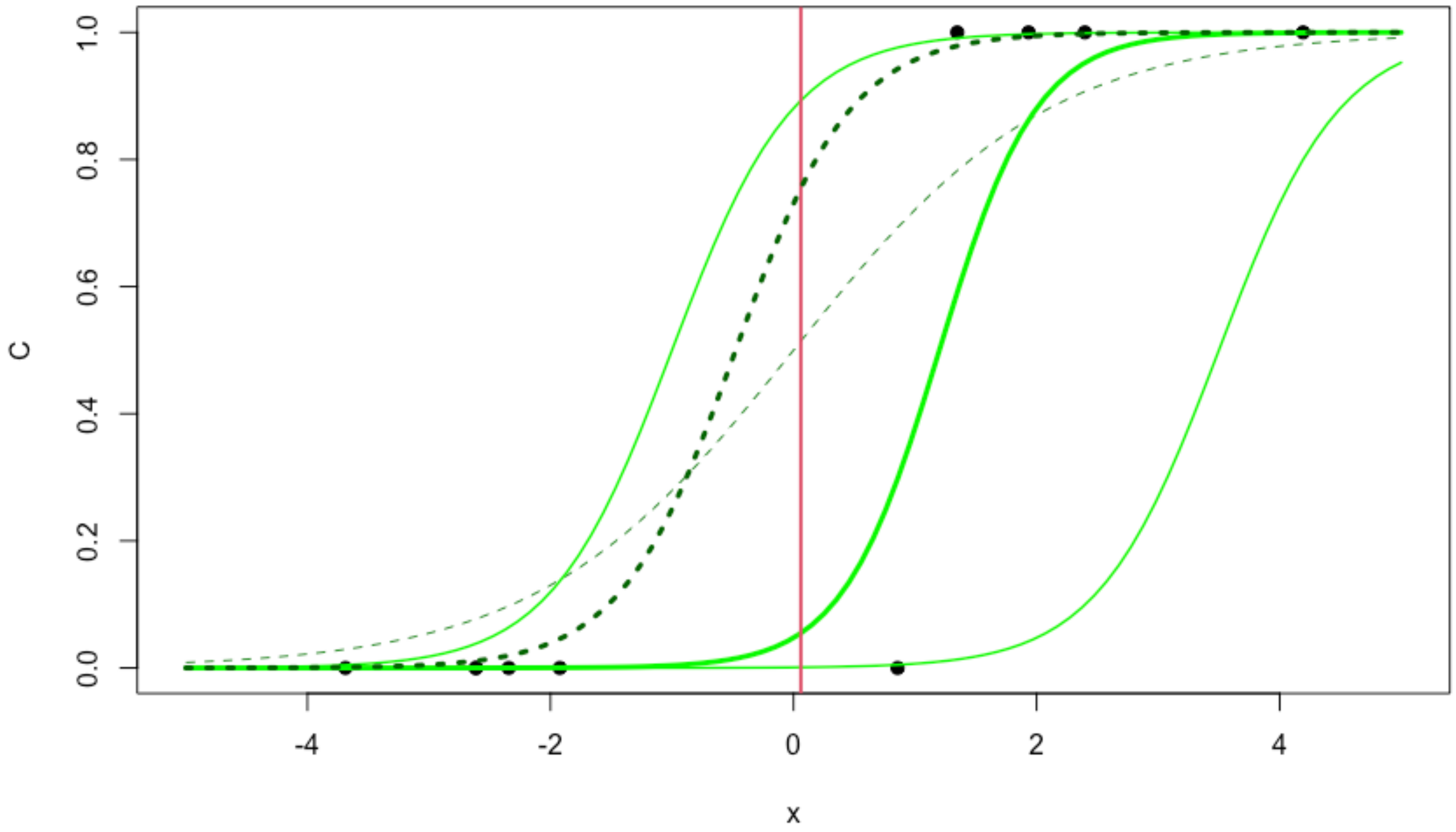
Thompson Sampling



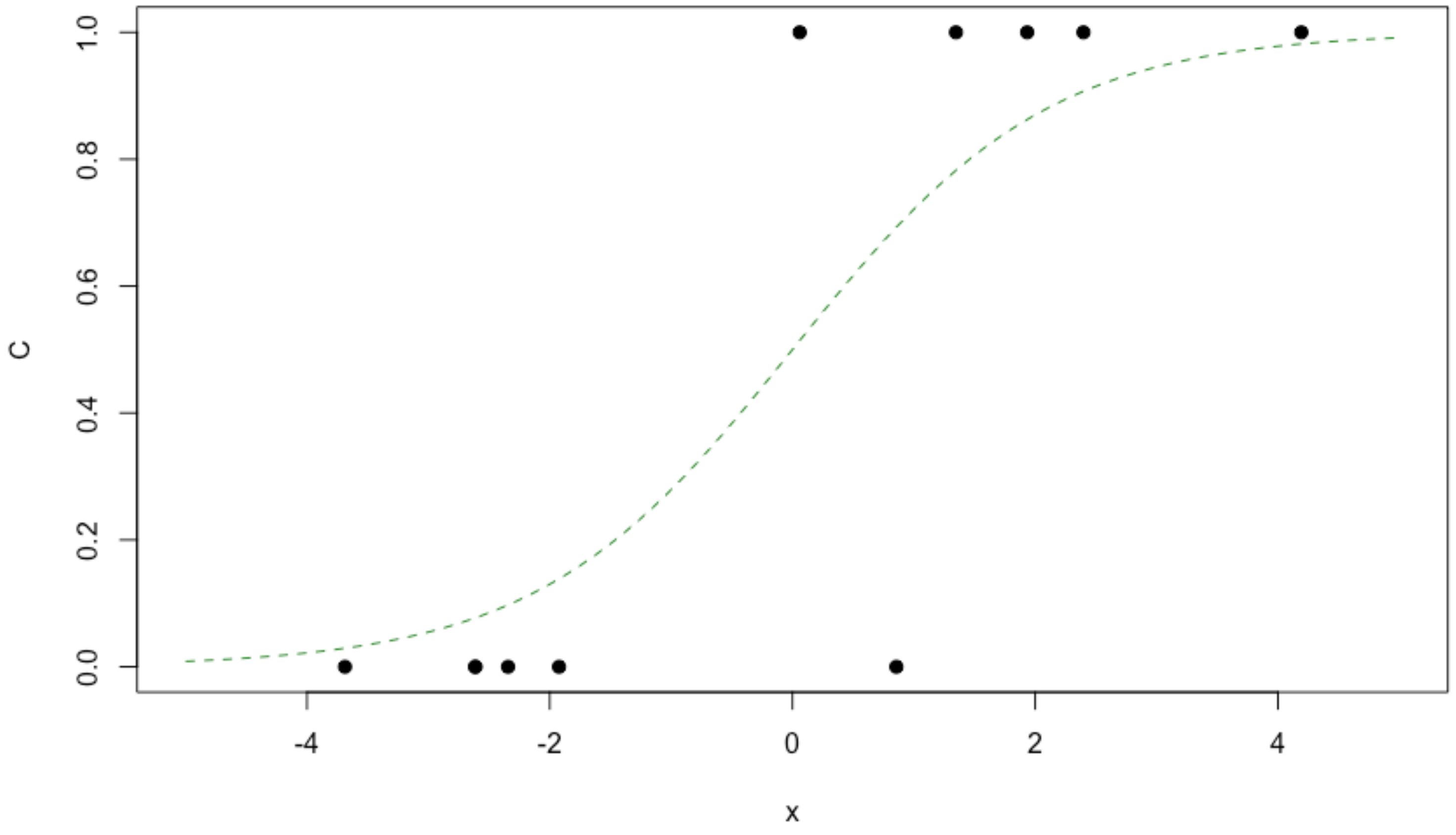
Thompson Sampling



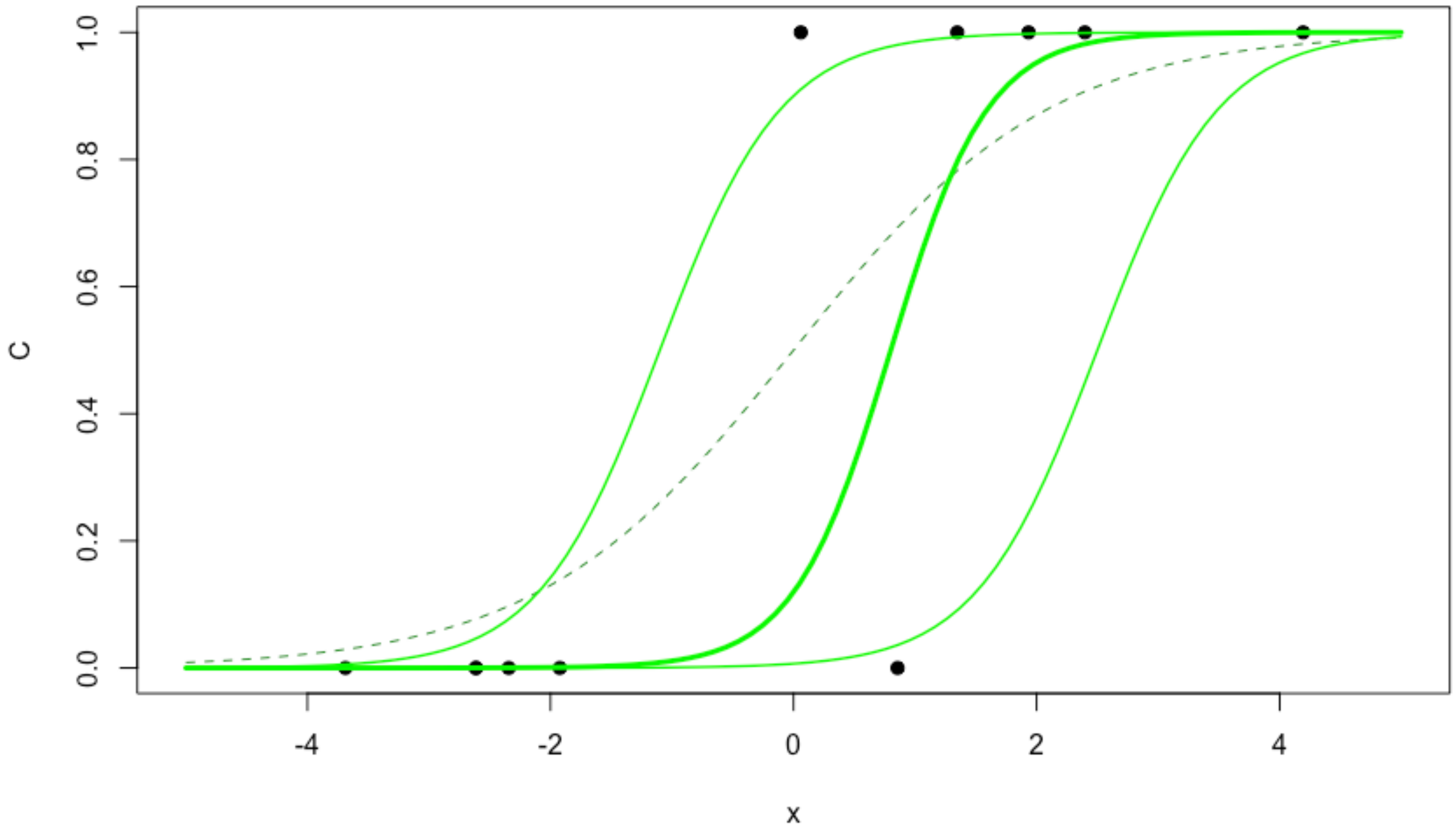
Thompson Sampling



Thompson Sampling



Thompson Sampling



Theoretical Aspects

Formalise the trade-off through Bayesian regret:

$$BReg(T) = E_{\pi_0}\left(\sum_{t=1}^{T}\left(\ell_0 \mathbb{I}\{C_t = 0, \tilde{C}_t = 1\} + \ell_1 \mathbb{I}\{C_t = 1, \tilde{C}_t = 0\}\right)\right),$$

where ℓ_0 and ℓ_1 are false positive and false negative costs resp.

Theoretical Aspects

Formalise the trade-off through Bayesian regret:

$$BReg(T) = E_{\pi_0}\left(\sum_{t=1}^{T}\left(\ell_0 \mathbb{I}\{C_t = 0, \tilde{C}_t = 1\} + \ell_1 \mathbb{I}\{C_t = 1, \tilde{C}_t = 0\}\right)\right).$$

Notoriously complex - generally intractable. Focus is on order results.

Theoretical Aspects

Formalise the trade-off through Bayesian regret:

$$BReg(T) = E_{\pi_0}\left(\sum_{t=1}^{T}\left(\ell_0 \mathbb{I}\{C_t = 0, \tilde{C}_t = 1\} + \ell_1 \mathbb{I}\{C_t = 1, \tilde{C}_t = 0\}\right)\right).$$

Notoriously complex - generally intractable. Focus is on order results.

An optimal algorithm will have Bayesian regret of order $O(\sqrt{dT})$ for any $\theta \in [0,1]^d$ (Bartok et al., 2014). We show optimality up to logarithmic terms for Thompson Sampling:

$$BReg^{TS}(T) = O\left(\sqrt{dT\log(T)}\right).$$

Theoretical Aspects

An optimal algorithm will have Bayesian regret of order $O(\sqrt{dT})$ for any $\theta \in [0,1]^d$ (Bartok et al., 2014). We show optimality up to logarithmic terms for Thompson Sampling:

$$BReg^{TS}(T) = O\left(\sqrt{dT\log(T)}\right).$$

This compares favourably to greedy heuristics ($O(T)$), and optimism based approaches (Bartok and Szepesvari, 2012):

$$BReg^{CBP-SIDE}(T) = O\left(d^2 \log(T)\sqrt{T}\right).$$

A (very sketchy) Proof Sketch

The expected regret in a single round t,

$$E_{\pi_0}\left(\ell_0 \mathbb{I}\left\{C_t = 0, \tilde{C}_t = 1\right\} + \ell_1 \mathbb{I}\{C_t = 1, \tilde{C}_t = 0\}\right).$$

Depends on probabilities of drawing a bad-sample,

$$P_{\pi_0}(\hat{C}\left(\tilde{\theta}_t\right) = 1 | C_t = 0) \text{ and } P_{\pi_0}(\hat{C}\left(\tilde{\theta}_t\right) = 0 | C_t = 1).$$

In turn, governed by expectation of $|x_t\tilde{\theta}_t - x_t\theta|$, (depends on π_t).

A (very sketchy) Proof Sketch

The expected regret in a single round t,

$$E_{\pi_0}\left(\ell_0 \mathbb{I}\left\{C_t = 0, \tilde{C}_t = 1\right\} + \ell_1 \mathbb{I}\{C_t = 1, \tilde{C}_t = 0\}\right).$$

Depends on probabilities of drawing a bad-sample,

$$P_{\pi_0}(\hat{C}\left(\tilde{\theta}_t\right) = 1 | C_t = 0) \text{ and } P_{\pi_0}(\hat{C}\left(\tilde{\theta}_t\right) = 0 | C_t = 1).$$

In turn, governed by expectation of $|x_t\tilde{\theta}_t - x_t\theta|$, (depends on π_t).

When π_t is well concentrated, few mistakes ($\sigma(x_t\theta) \approx \frac{\ell_0}{\ell_0+\ell_1}$).
When π_t is dispersed, tend to misclassify. If sufficiently many $C_t = 0$, then these errors bring information… and π_t concentrates.

A Caveat

- Draw a sample $\tilde{\theta}_t$ from the current posterior $\pi_{t-1}(\theta)$
- Treat $\tilde{\theta}_t$ as the true parameter and estimate $\hat{C}(\tilde{\theta}_t)$ based on x_t. Obtain an observation if $\hat{C}(\tilde{\theta}_t) = 1$, and update π.

A Caveat

- Draw a **sample** $\tilde{\theta}_t$ from the **current posterior** $\pi_{t-1}(\theta)$
- Treat $\tilde{\theta}_t$ as the true parameter and estimate $\hat{C}(\tilde{\theta}_t)$ based on x_t. Obtain an observation if $\hat{C}(\tilde{\theta}_t) = 1$, and **update** π.

A Caveat

- Draw a **sample** $\tilde{\theta}_t$ from the **current posterior** $\pi_{t-1}(\theta)$
- Treat $\tilde{\theta}_t$ as the true parameter and estimate $\hat{C}(\tilde{\theta}_t)$ based on x_t. Obtain an observation if $\hat{C}(\tilde{\theta}_t) = 1$, and **update** π.

- Bayesian inference for logistic regression is famously intractable.

- The previous theory depends* on $\tilde{\theta}_t$ being an exact sample from $\pi_{t-1}(\theta)$.

- When d is modest to large, rejection sampling can be highly inefficient, so we settle for an MCMC approximation.

Polya-Gamma Augmentation

- Posterior on θ in logistic regression is intractable:

$$\pi_t(\theta) \propto \pi(\theta) \prod_{i=1}^{t} \frac{\exp\left(x_i^T \theta\right)^{C_i}}{(1 + \exp(x_i^T \theta))}$$

- PG-Augmentation (Polson et al., 2012) adds Polya-Gamma latent variables.
  - Infinite sum of Gamma random variables, holding convenient identity for data augmentation.

- Admits a 2-stage Gibbs Sampler, where the PG-variables are sampled via rejection sampling with ≥ 0.9992 acceptance prob.

Polya-Gamma Thompson Sampling

Similar to Dumitrascu et al. (2018) we embed PG-Gibbs within Thompson Sampling.

- At time $t = 1, 2, \ldots$
  - Draw M samples $\{\tilde{\theta}_t^{\,m}\}_{m=1}^{M}$ via Gibbs Sampling initialised with $\tilde{\theta}_{t-1}^{M}$, targeting the current posterior $\pi_{t-1}(\theta)$
  - Treat $\tilde{\theta}_t^M$ as true parameter and estimate $\hat{C}(\tilde{\theta}_t^M)$ based on x_t.
  - If $\hat{C}\left(\tilde{\theta}_t^M\right) = 1$
    - Observe C_t and update the target distribution to $\pi_t(\theta)$.
  - If $\hat{C}\left(\tilde{\theta}_t^M\right) = 0$
    - Observe nothing and set $\pi_t(\theta) = \pi_{t-1}(\theta)$.

Polya-Gamma Thompson Sampling

Clearly M is important. As $M \to \infty$, this algorithm becomes equivalent to TS with exact sampling.

- At time $t = 1, 2, \ldots$
    - Draw M samples $\{\tilde{\theta}_t^{\,m}\}_{m=1}^{M}$ via Gibbs Sampling initialised with $\tilde{\theta}_{t-1}^{M}$, targeting the current posterior $\pi_{t-1}(\theta)$
    - Treat $\tilde{\theta}_t^M$ as true parameter and estimate $\hat{C}(\tilde{\theta}_t^M)$ based on x_t.
    - If $\hat{C}\left(\tilde{\theta}_t^M\right) = 1$
        - Observe C_t and update the target distribution to $\pi_t(\theta)$.
    - If $\hat{C}\left(\tilde{\theta}_t^M\right) = 0$
        - Observe nothing and set $\pi_t(\theta) = \pi_{t-1}(\theta)$.

Polya-Gamma Thompson Sampling

Clearly M is important. As $M \to \infty$, this algorithm becomes equivalent to TS with exact sampling.

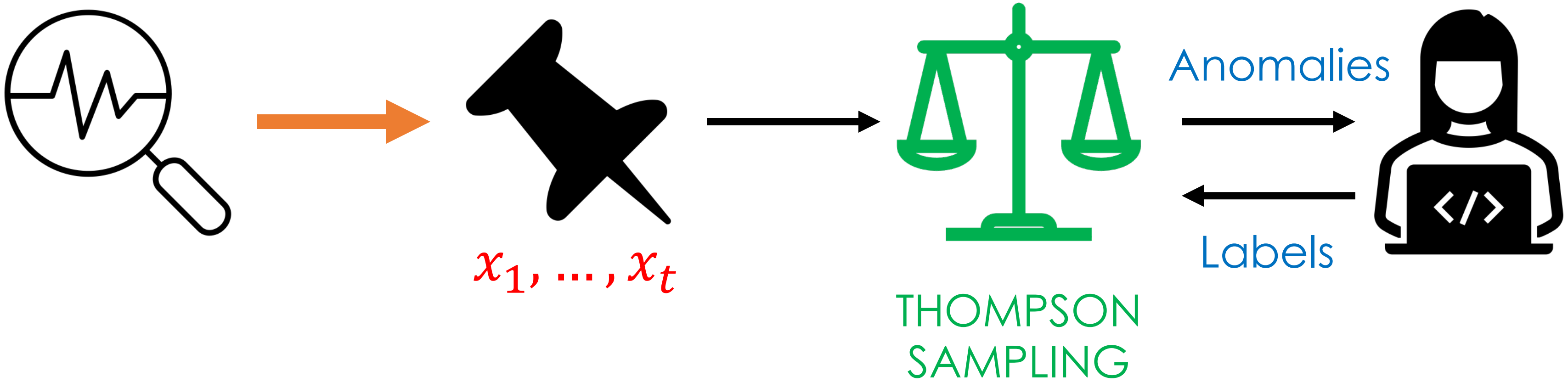
For finite M, the finite-time *BReg* guarantee does not extend*.

However,

- "Draw M samples $\{\tilde{\theta}_t^{\,m}\}_{m=1}^{M}$ via Gibbs Sampling initialised with $\tilde{\theta}_{t-1}^{M}$, targeting the current posterior $\pi_{t-1}(\theta)$."
- If $|\pi_t - \pi_{t-1}| \to 0$ as $t \to \infty$, then $\left\{\{\tilde{\theta}_t^{\,m}\}_{m=1}^{M}, \{\tilde{\theta}_{t+1}^{\,m}\}_{m=1}^{M}, \dots\right\}$ behaves like an infinite length chain in the limit.
- We can show asymptotic consistency of PGTS.

Summary

We've put anomaly detection and online classification (Apple Tasting via Thompson Sampling) together to produce a semi-autonomous algorithm.



Summary

We've put anomaly detection and online classification (Apple Tasting via Thompson Sampling) together to produce a semi-autonomous algorithm.

The approach allows us to **automate where possible**, without large amounts of initial labelled data, and continues to **learn as it proceeds**.

We have a theoretical guarantee* on the Bayesian regret.

Open Problems

There is a utility to utilising approximate sampling within TS.

Open Problems

There is a utility to utilising approximate sampling within TS.

How do we choose M to balance practical (computational cost) and theoretical (regret guarantees) aspects?

Open Problems

There is a utility to utilising approximate sampling within TS.

How do we choose M to balance practical (computational cost) and theoretical (regret guarantees) aspects?

When is a costly rejection sampler better?

Open Problems

There is a utility to utilising approximate sampling within TS.

How do we choose M to balance practical (computational cost) and theoretical (regret guarantees) aspects?

When is a costly rejection sampler better?

Do we need to use the exact posterior at all?

References & Contact

- Bartok, G., Foster, D.P., Pal, D. Rahklin, A., and Szepesvari, C. (2014). *Partial Monitoring – Classification, Regret Bounds and Algorithms.* Maths of OR.
- Bartok, G., and Szepesvari, C. (2012). *Partial Monitoring with Side Information.* ALT.
- Dumitrascu, B., Feng, K., and Engelhardt, B. (2018). *PG-TS: Improved Thompson Sampling for Logistic Contextual Bandits.* NeurIPS.
- Grant, J.A., Leslie D.S. (2021). *Apple Tasting Revisited: Partially Monitored Online Binary Classification*. In Submission, arXiv:2109.14412.
- Helmbold, D.P., Littlestone, M., and Long, P.M. (2000). *Apple Tasting*. Information and Computation.
- Polson, N.G., Scott, J.G, and Windle, J. (2013). *Bayesian Inference for Logistic Models using Polya-Gamma Latent Variables.* JASA.

j.grant@lancaster.ac.uk @james_a_grant