

Data Science Toolbox Question Sheet

12.1 Ethics and Privacy in Data Science

Daniel Lawson

Block 12

1. What is the single most important piece of legislation affecting data science in the EU?
2. How does the Human Rights Act affect data science in the UK?
3. Why is anonymity important? Why is it insufficient to simply remove identifiers when anonymising a dataset?
4. What is a statistical disclosure attack? Under what circumstances would a database be vulnerable to one? Who would the attacker be in that context?
5. Give an example of a statistical disclosure attack.
6. Given the definition of epsilon-differentially private, interpret what this means. What does the data controller have control over in this equation?
7. What are the differential privacy properties of the following algorithm in terms of attribute disclosure, in response to a question with a binary answer? Plug in the answer for $p = 0.5$ and $p = 0.01$.
 1. With probability p , respond truthfully.
 2. Otherwise, with probability 0.5 respond “yes” and otherwise respond “no”.
8. How can statistical disclosure attacks be mitigated? What are the trade-offs involved?
9. What is interpretability in the context of data science? Give an example of how different types of algorithm can be investigated to give an individual a reason why a decision was made about them.
10. What is algorithmic fairness? Why is it important?
11. Give three things a data scientist can do to address whether their algorithm is fair.
12. What is the difference between equality of outcome and equality of opportunity?