Grading rubric Causal Data Science 2022-2023 exam

General rule:

- If correct results, but no explanation or wrong explanation, half the points
- If correct explanation but wrong result, half the points

Multiple choice questions rubric:

Many exercises are multiple choice questions. If there is only one correct answer (circles), it gets all of the exercise points. If there are multiple possible correct answers (boxes), the total points of the exercise are split among the correct answers. For example if the exercise has 2 correct answers and the total points are 2, then each correct answer is assigned 1 point. If you only choose one of the two correct answers, you would get only 1 point.

Week 1 short questions

1a (1 point) - No OR We cannot say
1b (1 point) - Confounder
1c (1 point) - a. Simpsons' paradox
1d (1 point) - a. observational
1e (1 point) - a. counterfactual
1f (1 point) - b. Y is independent of X given Z

1g Factorizations (5 points)

- 1 point per UNIQUE factorization (there are exactly 5 of them)
    - $p(x)\ p(y|x)\ p(z|x)$
    - $p(y)\ p(x|y)\ p(z|x)$
    - $p(y)\ p(z|y)\ p(x|y,z)$
    - $p(z)\ p(x|z)\ p(y|x)$
    - $p(z)\ p(y|z)\ p(x|y,z)$

Week 2 short questions

2a (1 point) - j
2b (1 point) - neighbours OR adjacent nodes

2c (1 point) - b. No

2d D-separations (10 points)

- 1 point per correct d-separation (there are exactly 5 of them)
    - 0 _||_ 2
    - 0 _||_ 1
    - 0 _||_ 1 |2
    - 0 _||_ 2 |1
    - 2 _||_ 3 |0,4,1
- 1 point per correct explanation of each d-separation (e.g. showing all paths)

2e: Interventional distributions (5 points)

- 1 points for P(X) = N(0,5)
    a. 0.5 point for Mean(X) = - Mean(Y) - Mean(EpsX) = -0-0=0
    b. 0.5 point for Var(X) = (-1)^2 Var (Y) + (-1)^2 Var(EpsX) = Var (Y) + 1 = 2^2 Var[EpsY] + 1 = 4*1+1=5
- 1 points for P(X|do(Z=1) = P(X)
- 1 points for P(X|do(Y=1)) = N(-1, 1)
    a. 0.5 point for Mean (X|do(Y=1) = - Mean(Y|do(Y=1)) - Mean(EpsX) = -1 - 0 = -1
    b. 0.5 point for Var(X|do(Y=1) = (-1)^2 Var (Y|do(Y=1)) + (-1)^2 Var(EpsX) = 0 + 1 = 1
- 2 points for P(X|do(Y=3+EpsY)) = N(-3,2)
    a. 1 point for Mean(X|do(Y=3+EpsY)) = - Mean(Y|do(Y=3+EpsY)) - Mean(EpsX)= - Mean(3) - Mean(EpsY) + Mean(EpsX) = -3-0-0=-3
        i. 0.5 points for correct number without derivation
    b. 1 point for Var(X|do(Y=3+EpsY)) = (-1)^2 Var (Y|do(Y=3+EpsY)) + (-1)^2 Var(EpsX) = Var(3+EpsY) + 1 = Var(3)+Var(EpsY)+1 = 0+1+1=2
        i. 0.5 points for correct number without derivation

Week 3 short questions

3a (1 point) - adjustment criterion OR do-calculus (although it does more than valid adjustment sets)
3b (1 point) - mediator
3c (1 point) - Two boxes need to be crossed: 1. Z does not contain any descendant of the treatment i AND 3. Z blocks all backdoor paths from treatment i to outcome j
3d (1 point) - b. I causes the treatment

3e Backdoor adjustment (5 points)

- 0.5 point per each of 5 adjustment sets (there are 6 adjustment sets possible):

    - {0, 1}, {2}, {1, 2}, {0, 1, 2}, {0, 2}, {0}
- 0.5 point per explanation

3f: backdoor, adjustment, frontdoor (5 points)

- 1 point for mentioning the different results for **backdoor and adjustment** criterion (some vs ALL adjustment sets)
- 1 point for describing precisely the difference between **backdoor and adjustment criterion,** e.g. with an example graph or by describing the graphical criteria
- 1 point for mentioning the benefits of the **frontdoor criterion,** specifically that it can be used when some adjustment variable is **unobserved**
- 1 point for showing a causal graph or describing a case **where the frontdoor criterion applies and backdoor/adjustment criteria do not**
- 1 point for showing a causal graph or describing a case where **backdoor/adjustment criterion apply and frontdoor does not**

Week 4 short questions

4a (1 point) - ATC formula from slides $\mathbb{E}[Y(t = 1) - Y(t = 0) \mid T = 0]$
4b (1 point) - a. The treatment assignment of a unit i should not affect the treatment assignment of any other unit.
4c (1 point) - a. Reweigh the outcome of each unit by the inverse of the probability that they were assigned to whichever treatment option they were actually assigned to.

4d: Counterfactuals (5 points)

- Abduction - 1 point
- Prediction 1 - 2 points (with calculation)
- Prediction 2 - 2 points (with calculation)

See explanation here.

4e: exact matching, IPW (5 points)

- 1 point: Describe correctly exact matching
- 2 point: Describe correctly IPW (1 point simple explanation, 2 points more sophisticated)
- 2 point: Describe correctly that exact matching throws away data, while IPW does not, it only reweights them, so IPW is better when we don't have a lot of data

Week 5 short questions

5a (1 point) - CPDAG (acceptable although slightly wrong also MEC)

5b SGS exercise (5 points)

- 1 point skeleton, 0.5 partially correct
- 7 unshielded triples of which 2 v-structures, 0.5 points for each triple = 3.5 points
- 0.5 points Meek's rules
- In case misunderstandings on the theory half the points of each phase that is wrong

See explanation here.

5c (2 points): True

5d: GES exercise (10 points)

- 1 point for not violating DAGs/ in given CPDAG or making them clear
- 1 point for finding all four possible edges to be added
- 2*4 = 8 computations - 1 point each correct

See explanation here.

5e: constraint-based, restricted models (10 points)

- 2 points: Discuss the benefits of constraint-based methods, e.g., no need for parametric assumptions in the method, at most in the tests, some more advanced methods like FCI can also deal with latent confounding
- 2 points: Discuss the assumptions and drawbacks of constraint-based methods, e.g. we can only identify the causal graph up to an equivalence class, independence tests might be wrong, etc, as well as their benefits (no need for parametric assumptions in the method, at most in the tests)
- 2 point: Discuss the benefits restricted models, e.g. you can fully identify the causal DAG (not just the CPDAG)
- 4 points: Discuss the assumptions and drawbacks restricted models
    - 2 point: discuss additive noise models for ANM/RESIT - additive noise, independent noise variables, the two variables cannot be both Gaussian. Potentially mention example of when noise is not additive
    - 2 point: discuss linear non-gaussianity for LINGAM - linear SCM, non-Gaussian noise (potentially mention why, i.e. because it uses ICA), mention it fails if noises are Gaussian

Week 6 short questions

6a (1 point) - parents /direct causes
6b (2 points) - Yes

6c (3 points) - c. Neither P(X1) nor P(X1|X3)

6d: causality helps ML (10 points)

5 points - no explanation
6 points - a superficial explanation, e.g. mention in a superficial way a simple use cases, e.g. fairness, robustness, transfer learning, RL or the fact that most ML uses correlations
7 points - an adequate explanation, mention multiple use cases from class or provide more information on one use-case
8 points - a good explanation showing some theoretical understanding
9 points - a clear explanation with good theoretical understanding, possibly with some mention of the more specific applications discussed in class, e.g. how one can use d-separation to reason in transfer learning/domain adaptation, how one can learn causal graphs from multiple environments (e.g. JCI or ICP), how one can use causality in RL
10 points - a detailed explanation with strong theoretical understanding, describing several of the more specific applications discussed in class