Holistic AI × The Alan Turing Institute

Bias in Regression Tasks – Part I

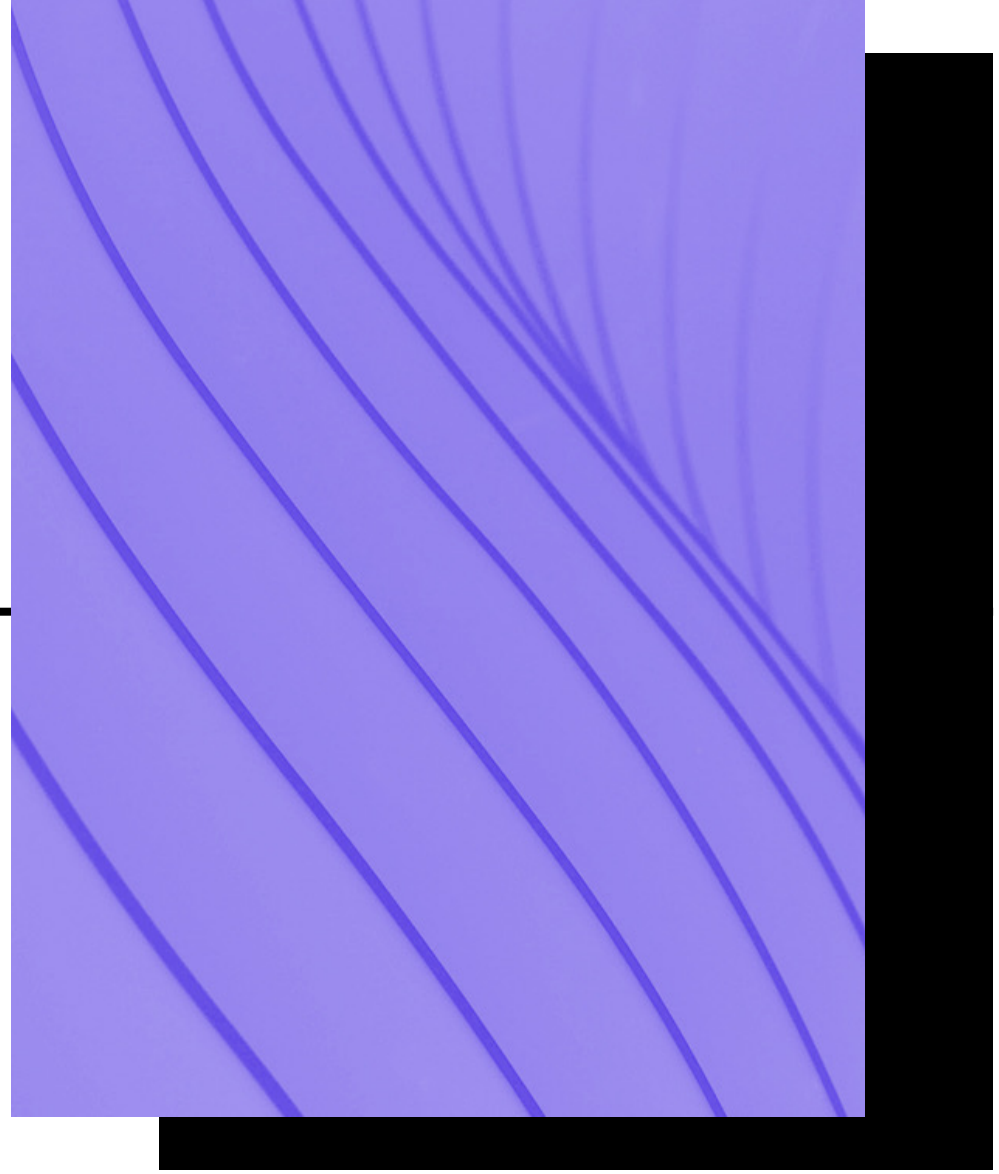
Content by: Sachin Beepath, Giulio Filippi, Nigel Kingsman, Cristian Munoz, Roseline Polle, Sara Zannone

Speaker: Sara Zannone

Contents

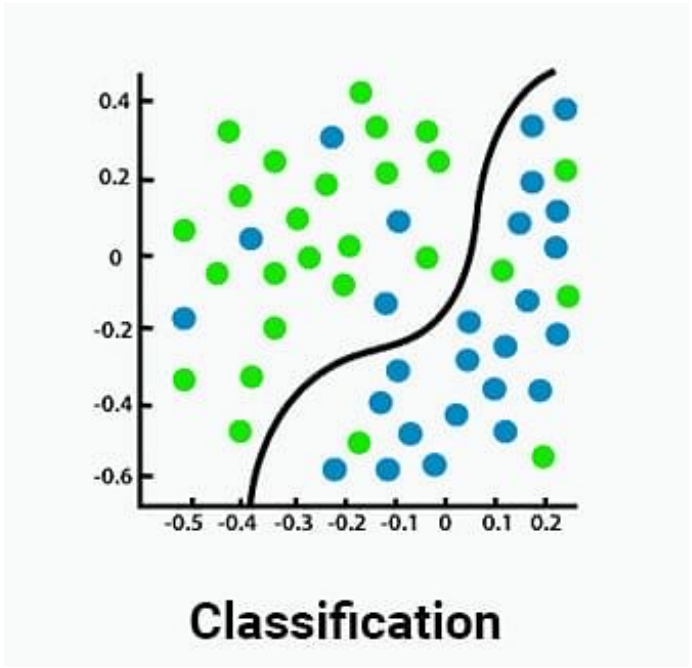
I - Introduction

- 1) Introduce Regression as a form of AI.
- 2) Provide real world examples to contextualize the ideas.
- 3) Motivate the importance of fairness for Regression.



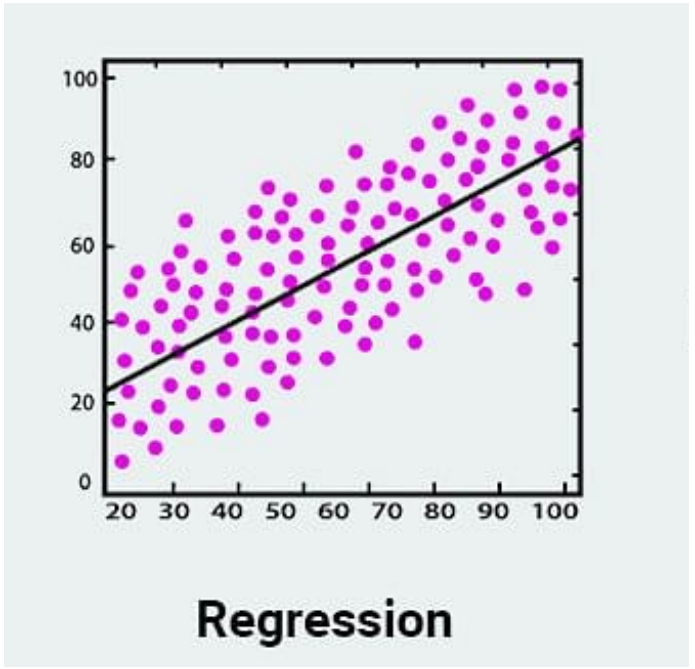
Classification tasks

- A machine learning algorithm learns to approximate the mapping from the input X to the output Y.
- In classification, the output Y is a binary variable
- The algorithm therefore learns to split the data in two classes: 0 or 1



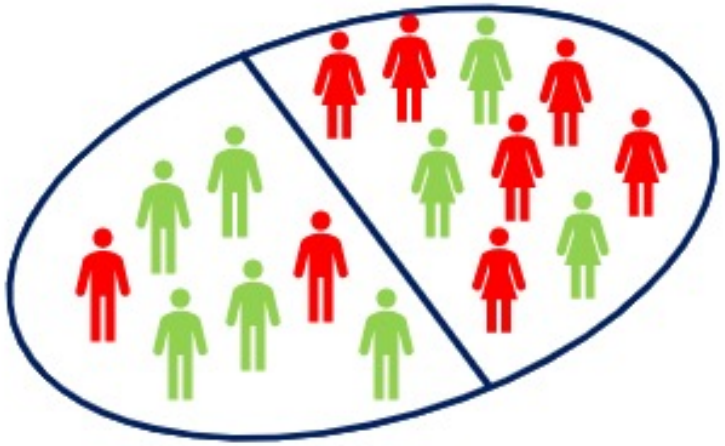
Regression tasks

- A machine learning algorithm learns to approximate the mapping from the input X to the output Y.
- In regression, the output Y is a continuous variable
- The algorithm therefore learns a function f(X) that can approximate the trend of the data



Example 1 - Recruitment

- **Binary classification**: an AI system is used to predict whether a candidate should be hired
- Model's outputs are:
  - 1 (pass)
  - 0 (fail)

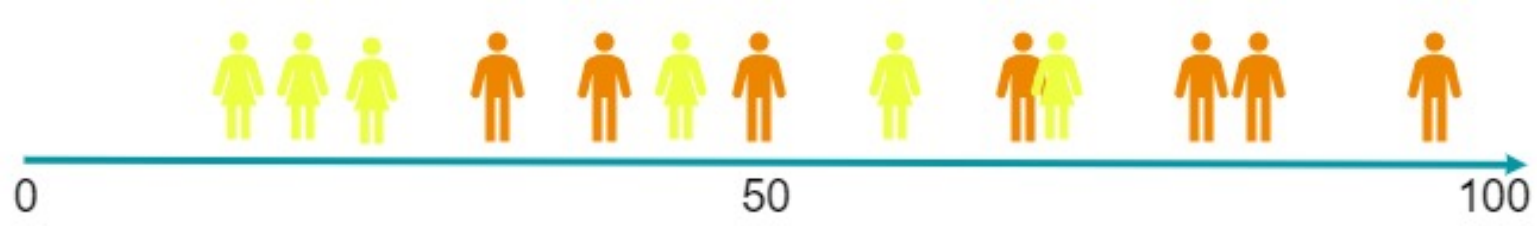


- The use of AI in recruitment is highly regulated (NYC Bias Audit Law)
  - Are the outcomes of the algorithm are fair for all?
    - Gender
    - Ethnicity

- We have covered how to measure and mitigate bias for this example in our previous course

Example 1 - Recruitment

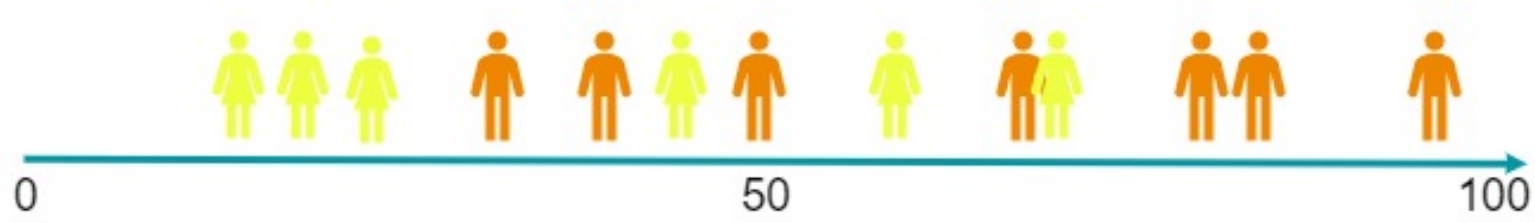
- **Regression**: The CVs of candidates are scored from 1 to 100



- This allows, for example, more autonomy for decision-makers (Veale et al., 2018)

Example 1 - Recruitment

- **Regression**: The CVs of candidates are scored from 1 to 100



- Similar to the high-profile case of the Amazon Recruitment Tool

- The NYC Bias Audit Law requires companies to calculate bias metrics for recruitment both in classification and regression tasks

Example 2 – Student grades

- Similarly to our previous example, AI technology can be used to predict students' average grades. Since the task would be to predict a continuous number, this would be a regression task

- In this case, it is important to make sure that the predicted grades are fair for all groups (gender, ethnicity etc…)

- Furthermore, they should not depend on other unfair criteria, for example, on the neighbourhood or the quality of the school

Example 2 – Student grades

- Similarly to our previous example, AI technology can be used to predict students' average grades. Since the task would be to predict a continuous number, this would be a regression task

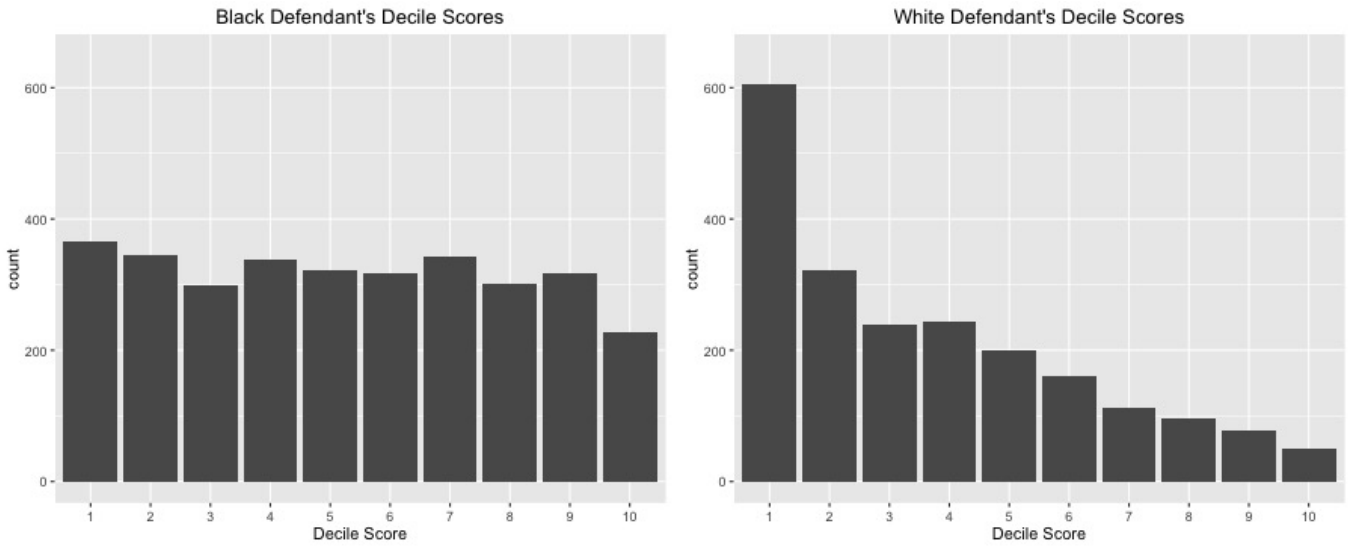
- In this case, it is important to make sure that the predicted grades are fair for all groups (gender, ethnicity etc…)

- Furthermore, they should not depend on other unfair criteria, for example, on the neighbourhood or the quality of the school

- high-profile case: [2020 UK school exam grading controversy]

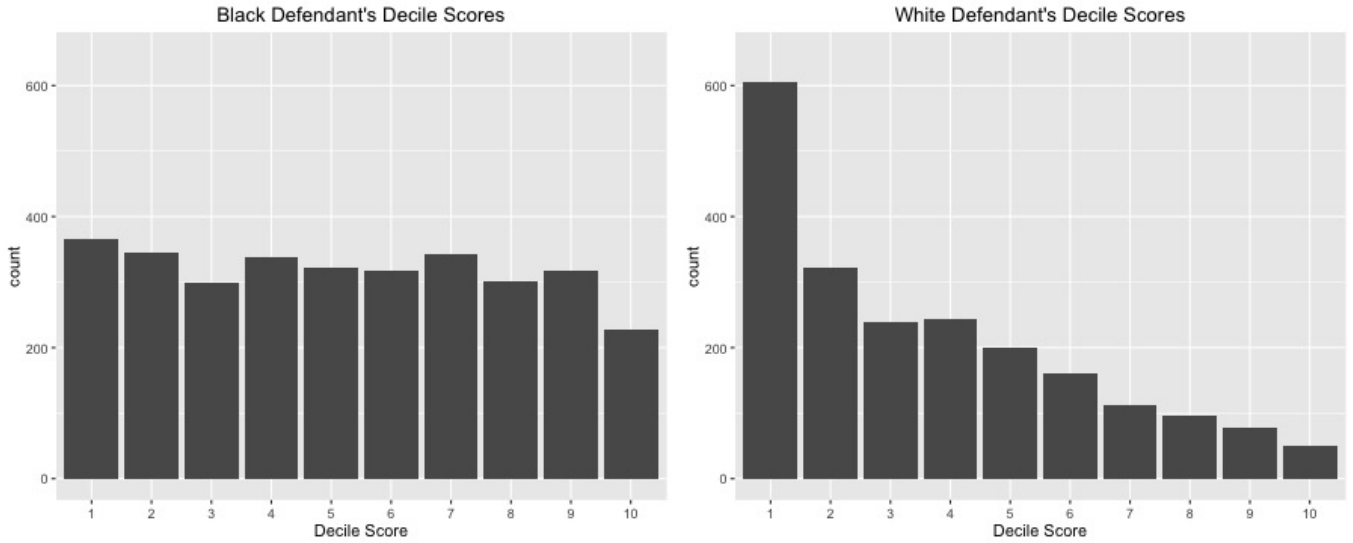
Example 3 – Crime Recidivism Prediction

- Courts have been using AI systems to predict the likelihood that a criminal defendant will commit another crime
- We should ensure that the system's error does not depend on protected attributes (e.g. ethnicity)



Example 3 – Crime Recidivism Prediction

- Courts have been using AI systems to predict the likelihood that a criminal defendant will commit another crime
- We should ensure that the system's error does not depend on protected attributes (e.g. ethnicity)



- High profile case: [COMPAS recidivism algorithm]
  - ranks prisoners' recidivism risk from 1 to 10
  - Black defendants were often predicted to be at a higher risk of recidivism than they actually were
  - White defendants were often predicted to be less risky than they were.

Example 4 – Credit scoring

- AI technology is also used in predicting credit scoring
- Sensitive application, which can have a huge impact on someone's life
- Credit scoring should not depend on protected attributes like gender and ethnicity
- High profile case: Apple Card gave David Heinemeier Hansson (tech entrepreneur) 20 times the credit limit that his wife got.

Conclusion

- Regression as a form of AI
- Real world example of regression tasks and the need for fairness
- Famous case studies of fairness in Regression

Contents