

✔ Congratulations! You passed!

Grade received 83.33% To pass 80% or higher

Go to next item

Model Analysis and Debugging

Total points 6

1. When evaluating an ML model during training the goal is to improve top-level metrics such as overall accuracy. This information is used to decide whether the model is doing well or not, but it doesn't show how well it does on individual parts of the data. Which technique is extremely helpful to address this shortcoming?

1 / 1 point

- ☐ Streaming metrics
- ☐ TensorFlow Metric Analysis (TFMA)
- ☒ Data Slicing
- ☐ Apache Beam

✔ Correct

That's right! Slicing deals with understanding how a model is performing on each subset of data.

2. Streaming metrics are approximations to full-pass performance metrics computed on _____.

1 / 1 point

- ☐ the full data set
- ☐ the full validation data set.
- ☒ mini-batches of data
- ☐ slices of data

✔ Correct

That's right! This is a nice way to approximate the full-pass metrics without incurring a huge computational overhead cost.

3. A recent credit card loyalty program offered by a big technology company has been labeled as “sexist”, a clear example of algorithm based social discrimination. Let's examine a user complaint on Twitter: “My wife and I filed joint tax returns, live in a community-property state, and have been married for a long time. Yet the black box algorithm thinks I deserve 20x the credit limit she does. No appeals work.” These and other similar claims have triggered a full-blown investigation by the New York State Department of Financial Services. Which of the reviewed techniques in lecture could have been implemented to prevent this embarrassing problem?

1 / 1 point

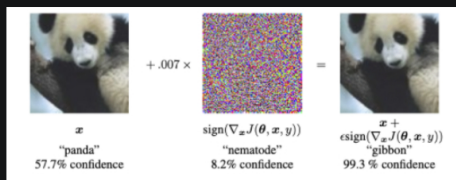
- ☐ Residual analysis
- ☐ Model robustness
- ☐ Data slicing
- ☒ Model debugging

✔ Correct

That's right! Model debugging tries to improve the transparency of models by highlighting how data is flowing inside and thus can prevent harmful social discrimination.

4. State of the art convolutional neural networks can be fooled to misclassify craftily noise corrupted images with changes that are completely Imperceptible to the human eye, as illustrated by the following picture:

1 / 1 point



What type of analysis can help us detect and prevent these types of scenarios?

- ☒ Sensitivity analysis
- ☐ Residual Analysis
- ☐ Adversarial attack
- ☐ Dimensionality reduction

✔ **Correct**

That's right! Sensitivity analysis helps with understanding a model by examining the impact that each feature has on the model's prediction. In sensitivity analysis we experiment by changing a feature value while holding the other features constant, and observe the model results.

5. A performance-metric gap between two or more groups could be a sign that an ML model may have unfair skews. Therefore, is achieving performance equality (on fairness indicators) across groups a definite sign that a model is fair?

1 / 1 point

☐ Yes

☒ No

✔ **Correct**

That's right! Systems are highly complex and achieving equality on one, or even all of the provided metrics can't guarantee fairness. Fairness evaluations should be run throughout the development process and post-launch as well.

6. After a model has been deployed, is it usually feasible to perform residual analysis?

0 / 1 point

☐ No

☒ Yes

✘ **Incorrect**

Not quite! You will need to implement an infrastructure to label new data as it reaches your production system.