Try again once you are ready

Grade received 50% To pass 80% or higher

Try again

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
- ◉ Data Slicing
- ○ TensorFlow Metric Analysis (TFMA)
- ○ 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 _____. **0 / 1 point**

- ○ mini-batches of data
- ◉ slices of data
- ○ the full validation data set.
- ○ the full data set

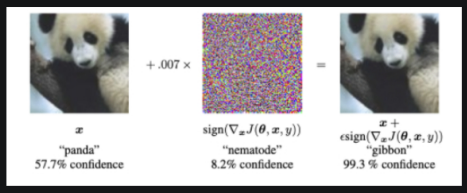
Incorrect
Not quite! While slicing helps reduce the batch size over which you compute performance, it biases the sample towards your selection criteria.

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? **0 / 1 point**

- ○ Model robustness
- ○ Data slicing
- ○ Model debugging
- ◉ Residual analysis

Incorrect
Not quite! Residual analysis is a subset of a larger class of approaches to prevent social discrimination and other problems that ML prodcutioon systems usually face.

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: **0 / 1 point**



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

- ○ Sensitivity analysis
- ○ Dimensionality reduction
- ○ Residual Analysis
- ◉ Adversarial attack

Adversarial attack

Incorrect
Not quite! The picture shown is an example of adversarial attack. What type of analysis can be done to prevent/understand this and other attacks?

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? 1 / 1 point

- Yes
- No

Correct
That's right! Once your model is deployed, you may not have a good amount of labeled data and consequently, residual analysis can prove to be a costly exercise as it might include you hiring workers to label your incoming test data.