Privacy Accounting and Quality Control in Sage
Why is DP needed with ML?

- ML datasets could leak specifics about individual entries in their training sets.
- Prevent featurization of dataset
  - Membership inference
  - Reconstruction attacks
Q: Why can’t you just train a ML model using PINQ?
Sage Access Control & privacy adaptive training

Leverages the idea that the growing database is not static but growing, keeps training models endlessly on sensitive data stream.
Challenges

Privacy Utility trade-off:

- Less accurate results that fail to meet the quality targets more often than w/o DP.
- Low-quality models whose validations succeed by chance.
Splitting the data

- **User-Level:** based on user ID
  - Use incrementing userID’s, max stored
  - New blocks are only created when new users join

- **Event-level:** splitting on time
  - days, months, etc.
Taxi Example

- Preprocessing_fn: makes aggregate features i.e distance of ride, hour of day
  - Dp_group_by_mean:
    - Number of times key appears
    - Sum of values associated w/ key
  - Each data point has one key

```python
def preprocessing_fn(inputs, epsilon):
    dist_01 = tft.scale_to_0_1(inputs["distance"], 0, 100)
    speed_01 = tft.scale_to_0_1(inputs["speed"], 0, 100)
    hour_of_day_speed = group_by_mean_age.
dp_group_by_mean(inputs["hour_of_day"], speed_01, 24, epsilon, 1.0)
    return ("dist_scaled": dist_01,
    "hour_of_day": inputs["hour_of_day"],
    "hour_of_day_speed": hour_of_day_speed,
    "duration": inputs["duration"])

def trainer_fn(hparams, schema, epsilon, delta): [...] feature_columns = [numeric_column("dist_scaled"),
    numeric_column("hour_of_day_speed"),
    categorical_column("hour_of_day", num_buckets=24)] estimator = tf.estimator.DPNRegressorSage.DPNNRegressor(config=run_config,
    feature_columns=feature_columns,
    dnn_hidden_units=hparams.hidden_units,
    privacy_budget=(epsilon, delta))
    return tfx.executors.TrainingSpec(estimator,...)

def validator_fn(epsilon):
    model_validator = tfx.components.ModelValidatorSage.DPModelValidator(
        examples=examples_gen.outputs.output,
        model=trainer.outputs.output,
        metric_fn=_MSE_FN, target = _MSE_TARGET,
        epsilon=epsilon, confidence=0.95, B=1)
    return model_validator

def dp_group_by_mean(key_tensor, value_tensor, nkeys,
    epsilon, value_range):
    key_tensor = tf.dtypes.cast(key_tensor, tf.int64)
    ones = tf.fill(tf.shape(key_tensor), 1.0)
    dp_counts = group_by_sum(key_tensor, ones, nkeys) + laplace(0.0, 2/epsilon, nkeys)
    dp_sums = group_by_sum(key_tensor, value_tensor, nkeys) + laplace(0.0, value_range + 2/epsilon, nkeys)
    return tf.gather(dp_sums/dp_counts, key_tensor)
```
Sage Access Control: requirements for composition theory

- R1: Multiple training pipelines w/ differing amounts of data needed for performance
- R2: Adaptivity in choice of queries, DP parameters and data subsets
- R3: Some models are ran periodically w/ new data and others are retired
Failed Methods: which rules do these violate?

1. Query across the entire stream:
   - $\epsilon_D = \epsilon_1 + \epsilon_2 + \epsilon_3$

2. Queries split into subqueries and each run DP on individual blocks, results aggregated

3. A new data point is allocated to one of the waiting queries, which consumes entire privacy budget.
Block Composition Theory cont.

- Splits data into disjoint blocks adaptively chosen (R1, R2)
- Privacy loss of three queries will be max of $\epsilon_1 + \epsilon_2$, and $\epsilon_2 + \epsilon_3$
- New blocks D5 arrive with privacy loss of zero (R3)

System can run endlessly by training new models on new data!
Q: What does it mean for DP parameters to be chosen Adaptively?
Adaptive Parameters

Fig. 3. Characteristics of Data Interaction in ML.
Privacy-Adaptive Training

- To improve DP quality:
  - Increase privacy budget ($\epsilon, \delta$) or increase dataset size
- Accept: prediction target reached
- Retry: more data needed for assessment
- Reject: model will never reach target w/ sample size/privacy requirements
Discuss:

Q: What assumptions are made about the data? In what cases could Sage potentially not perform well?