Minimax Group Fairness: Algorithms and Experiments
Emily Diana, Wesley Gill, Michael Kearns, Krishnaram Kenthapadi, Aaron Roth
University of Pennsylvania, Amazon AWS AI

Motivation
- Machine learning researchers and practitioners have often focused on achieving fairness with respect to protected attributes (race, gender, ethnicity, etc.)
- Equality of error rates is one of the most intuitive and well-studied group fairness notions
- But in practice, equalizing error rates and similar notions may require artificially inflating error on easier-to-predict groups and may be undesirable for a variety of reasons
- There are many social applications of machine learning in which most/all of the targeted populations is disadvantaged
- Might be interested in ensuring predictions are roughly equally accurate across racial groups, income levels, geographic location, etc
- But, if this can only be achieved by raising lower group error rates, then we have worsened overall social welfare
- Therefore, might be preferable to consider the alternative fairness criterion of minimax group error, recently proposed by [Martinez, 2020]

Contributions
1. Propose two algorithms, both two player zero-sum games:
   1. **INIMAX**: Finds a minimax group fair model from a given statistical class
   2. **INIMAXRELAXED**: Finds a model that minimizes overall error subject to the constraint that all group errors must be below a predetermined threshold
   - NAVIGATES tradeoffs between notions of fairness and overall accuracy
2. Prove that both algorithms converge and are oracle efficient.
3. Show how our framework can be extended to handle different types of error rates, such as equal errors classification
4. Provide a thorough experimental analysis of our two algorithms under different settings

Mathematical Framework
- Consider pairs of dependent and independent variables \((X, y)\)^{n_i}_{i=1} divided into \(K\) groups \((G_1, \ldots, G_K)\), class \(H\) of (potentially unfair) mixtures of statistical models, with loss function \(L\), and average group loss \(\omega_h\) for some \(h \in H\):
  \[
  \omega_h = \sum_{i=1}^{n_i} L(h(x), y)
  \]
- In pure minimax problem, goal is to find a mixed strategy \(h^*\) that minimizes the maximum error rate over all groups:
  \[
  h^* = \arg\min_h \max_{i=1}^{K} \omega_h(h)
  \]
- In relaxed version, specify max group error \(\gamma\) and model that minimizes overall population error while staying below the maximum group error threshold:
  \[
  \begin{align}
  \min_h & \omega_h(h) \\
  \text{subject to} & \omega_h(h) \leq \gamma, 0 < k \leq 1, \ldots, K
  \end{align}
  \]

Algorithmic Formulation: Two Player Zero-Sum Game
Can recast both problems as a zero-sum game between a (L)earner and a (R)egulator:
- At each round \(t\), there is a weighing over groups determined by \(R\)
- \(L\) (best) responds by computing model \(h_t\) to minimize the weighted prediction error
- \(R\) updates group weights using exponential weights/gradient ascent with respect to group errors achieved by \(h_t\)
- \(L\)s final model \(U\) is uniform distribution over all of \(h_t\)s produced

Minimax Fairness

Comparison of Minimax and Equal Errors Solutions on Seoul Bike Dataset

<table>
<thead>
<tr>
<th>Group</th>
<th>MINIMAX</th>
<th>Equal Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women</td>
<td>0.32</td>
<td>0.34</td>
</tr>
<tr>
<td>Men</td>
<td>0.33</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Dataset: Public bikes rented at each hour in Seoul Bike sharing system
Label: Rented bikes (normalized), Group: Season

Fairness Accuracy Tradeoff with MINIMAXRELAXED

Comparison of Minimax and Equal Errors on Marketing Dataset

<table>
<thead>
<tr>
<th>Group</th>
<th>MINIMAX</th>
<th>Equal Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>0.35</td>
<td>0.37</td>
</tr>
<tr>
<td>High</td>
<td>0.32</td>
<td>0.34</td>
</tr>
</tbody>
</table>

Dataset: Direct marketing campaigns (phone calls) of a Portuguese bank
Label: client subscribes term deposit, Group: Job

Generalization Results
- With probability \(1 - \delta\), generalization gap per group bounded by
  \[
  O \left( \log \frac{d}{\delta} \log \frac{n}{\delta} \right)
  \]
  where \(d\) is VC dimension of class \(H\), and \(n_i\) is sample size of group \(i\)
- Generalization gap for minimax group is bounded by
  \[
  O \left( \max \left\{ \log \frac{d}{\delta} \log \frac{n}{\delta} \right\} \right)
  \]
  i.e. dominated by sample size of the smallest group

Generalization Experiments

Minimax vs. Equal Errors Regression

Comparison of Minimax and Equal Errors Classification

Minimax vs. Equal Errors Classification

Dataset: Network connection data used to distinguish between 'bad' connections, called intrusions or attacks, and 'good' normal connections.
Label: Connection Legitimacy, Group: Protocol Type

Selected References
- [Martinez, S. Moro and P. Cortez and P. Rita (2014)]
- [Fairness Accuracy Tradeoff with M. Kenthapadi, K. Roth, A. Roth (2020)]
- [Minimax Group Fairness: Algorithms and Experiments, Martinez, M. Kenthapadi, K. Roth, A. Roth (2020)]
- [Minimax Group Fairness: Algorithms and Experiments, Martinez, M. Kenthapadi, K. Roth, A. Roth (2020)]
- [Minimax Group Fairness: Algorithms and Experiments, Martinez, M. Kenthapadi, K. Roth, A. Roth (2020)]