Agenda

- Introduction
- Model Development
- Example
- Conclusions
- Questions and Comments
Increasing expectation on financial institutions utilizing consumer data to “explain” how automated systems make decisions

ML models complicate explanations of “actionable” consumer behavior due to the inherent nonlinearity and interactions present in these models

Explanations of credit risk models do not necessarily translate into the ordered sequence of actions a consumer could take to improve their score to a desired value

**Objective**
Describe a method for constructing an optimal path that explicitly navigates an individual consumer through the model feature space from their current score to a score of their choosing
Explaining Credit Scores

Suppose a consumer desires to reach a given credit score…

What are the sequence of actions the consumer should take to reach the desired score?

“What If” Simulators

What would happen if an action is taken?

• What if I apply for a new credit card?
• What if I pay off my balance in full?

Reason Codes

Reasons why you were denied credit.

• Too many accounts past due
• Sum of balances on accounts too high
Explaining Credit Scores

“What If”

- Assumes independence of groups of attributes in the model and additivity
- No sense of “feasibility” to the consumer, only portfolio level
  - Can’t remove a bankruptcy received last month
  - Reduce $500K mortgage balance to $0

Reason codes (generic advice) provide factors most negatively impacting consumers score

- Max Points Lost
  Reason Code: $x_1$ $f[x_1^*, x_2^*; \beta] - f[x_1, x_2; \beta] = 0.30$
  Reason Code: $x_2$ $f[x_1^*, x_2^*; \beta] - f[x_1^*, x_2; \beta] = 0.29$

- Points to Max Improvement
  Reason Code: $x_1$ $f[x_1^*, x_2; \beta] - f[x_1, x_2; \beta] = 0.30$
  Reason Code: $x_2$ $f[x_1, x_2^*; \beta] - f[x_1, x_2; \beta] = 0.29$

Reason codes do not necessarily indicate the proper order of actions to improve score in the most efficient manner
Explaining Credit Scores

- How do attributes change across space and time?

- Every point in the scoring domain may not be feasible to the consumer

- Feasibility needs to account for covariance among attributes over time

- Given an initial starting point $x^0$ at $t^0$, we need to find $x^1 = x^0 + \delta$ at $t^1$ with the constraint that $f(x^1) = C$

Gradient descent (blue) and Mahalanobis (black) over time from current score ($x^0, t^0 =$ red dot) to specified score (yellow line) with a score increase of 0.025 units in each time period (orange points $= t^0$, blue points $= t^1$)*

*Score $= 1.0x_1 + 0.1x_2 + 1.0x_1x_2$
Seeking Feasibility

Need a metric that captures what movements in the feature space are feasible for a consumer over time, conditioned on their current location \( x_{t_0} \). Let,

\[
\delta = x_{t_1} - x_{t_0}, \quad \mu = \mathbb{E}(\delta|x_{t_0}), \quad \Sigma = \text{Cov}[\delta|x_{t_0}]
\]

- \( x_{t_0}, x_{t_1} = p \times 1 \) vectors of attributes measured at two time points
- \( \mu = p \times 1 \) vectors of mean differences of attributes at \( t_0 \) and \( t_1 \)
- \( \Sigma = \) covariance matrix of differences of attributes at \( t_0 \) and \( t_1 \)

Constrained Optimization Problem

argmin_{\delta} (\delta - \mu)^T \Sigma^{-1} (\delta - \mu)

Subject to: \( f(x_{t_0} + \delta; \beta) = C \)

Equation 1 finds the shortest Mahalanobis Distance (MD) between \( x_{t_0} \) and \( x_{t_1} = x_{t_0} + \delta \) to achieve a specified score increase.

MD accounts for different variances among attributes and covariance between attributes.
### Estimating Feasibility

- Need to estimate \( \mu \) and \( \Sigma \) across the domain of our feature space
- Using \( X^t \), partition the feature space into \( k \) clusters and estimate the mean \( \mu_k \) and covariance matrix \( \Sigma_k \) of the delta changes between time \( t_0 \) and \( t_1 \) per cluster
- Our objective function now becomes:

\[
\arg\min_\delta: (\delta - \mu_k)^T \Sigma_k^{-1} (\delta - \mu_k)
\]

Subject to:
\[
f(x^t_0 + \delta; \beta) = C
\]

**Equation 2 makes explicit that attributes can vary differently across time given a consumer’s current location in the domain.**
Optimal Path Algorithm

**Algorithm**

1. Conduct a cluster analysis on sample data at $t_0$
2. For each cluster, $c_k$, Calculate $\mu_k$, $\Sigma_k$
3. Given a consumer with attributes $x^{t_0}$:
4. for $j = 1, \ldots, T$ do
   a. Classify consumer with attributes $x^{t_{j-1}}$ into $c_k$
   b. $\arg\min (\delta - \mu_k)^T \Sigma_k^{-1} (\delta - \mu_k)$
      Subject to: $f(x^{t_{j-1}} + \delta; \beta) = C$
   c. Set $x^{t_j} = x^{t_{j-1}} + \delta$
5. end

Equation 3 requires mixed-integer nonlinear optimization

Bounds on attributes: (# Inquiries $\geq 0$), integer constraints (# inquires)

- Missing values
- Default values
- Categorical values
A Simple Use Case

Sampled 1% of U.S. consumers

Jan 2018 = \( t_0 \)
Feb 2018 = \( t_1 \)

- All attributes were treated as continuous (only 11 attributes were utilized in this use case)
- Default/Missing values imputed to 0 or 1 as appropriate
- Equifax NeuroDecision: 2-Hidden Layer (10-nodes each layer)
- K-means cluster analysis (100 clusters)
• **95% prediction intervals** \( (\delta = x^{t_1} - x^{t_0}) \) for two clusters (# Inquiries in 1- and 12-months)

• **Red**: No mean change in IQ1 and IQ12. Typical of consumers with no IQ on file or an inquiry between 2- and 11-months

• **Blue**: Mean change in IQ1 but not IQ12. Typical of consumers who have a recent inquiry that rolls off the IQ1 but not IQ12

---

• **95% prediction intervals** \( (\delta = x^{t_1} - x^{t_0}) \) for five clusters (balance and utilization)

• **Green/Blue**: Change in utilization, minimal change in balance. Typical of consumers with newly opened or closed accounts, credit limit increases or decreases, or low credit limits

• **Other colors**: Most clusters exhibit positive correlation: increasing balance increases utilization
Optimal Path For Consumer 1

Consumers goal is to **increase score 100-points** in 5-point increments every month

<table>
<thead>
<tr>
<th>Attribute</th>
<th>$t_0$</th>
<th>$t_{21}$</th>
</tr>
</thead>
<tbody>
<tr>
<td># Inquiries Last Month</td>
<td>0</td>
<td>0.67</td>
</tr>
<tr>
<td>Revolving Accts. Balance</td>
<td>$45,500$</td>
<td>$10,483$</td>
</tr>
<tr>
<td>Utilization Revolving Accts.</td>
<td>0.747</td>
<td>0.172</td>
</tr>
<tr>
<td>% Satisfactory Accts.</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td># Inquiries Last 12 Mos.</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td># Accts.</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td>Age Oldest Revolving Acct.</td>
<td>278</td>
<td>299</td>
</tr>
<tr>
<td># Accounts PD &gt; $0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total PD $</td>
<td>$0</td>
<td>$0</td>
</tr>
<tr>
<td># 60+ DPD Last 24 Mos.</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>% Revolving Accts. To Total Accts.</td>
<td>0.875</td>
<td>0.890</td>
</tr>
</tbody>
</table>

**SCORE**

<table>
<thead>
<tr>
<th>Score at $t_0$</th>
<th>Score at $t_{21}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>809</td>
<td>910</td>
</tr>
</tbody>
</table>
Consumers goal is to **increase score 100-points** in 5-point increments every month

<table>
<thead>
<tr>
<th>Attribute</th>
<th>$t_0$</th>
<th>$t_{21}$</th>
</tr>
</thead>
<tbody>
<tr>
<td># Inquiries Last Month</td>
<td>0</td>
<td>0.38</td>
</tr>
<tr>
<td>Revolving Accts. Balance</td>
<td>$542</td>
<td>$243</td>
</tr>
<tr>
<td>Utilization Revolving Accts.</td>
<td>1.0</td>
<td>0.48</td>
</tr>
<tr>
<td>% Satisfactory Accts.</td>
<td>0.12</td>
<td>0.38</td>
</tr>
<tr>
<td># Inquiries Last 12 Mos.</td>
<td>3</td>
<td>1.1</td>
</tr>
<tr>
<td># Accts.</td>
<td>8</td>
<td>7.6</td>
</tr>
<tr>
<td>Age Oldest Revolving Acct.</td>
<td>85</td>
<td>106</td>
</tr>
<tr>
<td># Accounts PD &gt; $0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Total PD $</td>
<td>$1421</td>
<td>$7.2</td>
</tr>
<tr>
<td># 60+ DPD Last 24 Mos.</td>
<td>10</td>
<td>1.25</td>
</tr>
<tr>
<td>% Revolving Accts. To Total Accts.</td>
<td>0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>

**SCORE**

<table>
<thead>
<tr>
<th></th>
<th>$t_0$</th>
<th>$t_{21}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>433</td>
<td>535</td>
</tr>
</tbody>
</table>
Conclusions

Most credit score simulators are limited in available actions and are “trial and error” based.

Reason codes do not necessarily indicate the proper order of actions to improve a consumers score in the most efficient manner.

Our algorithm generates an optimal path that is feasible for a consumer in monthly increments:
- Our algorithm accounts for within-subject changes in attributes over space and time.

The optimal paths algorithm provides feasible, actionable, and impactful recommendations to the consumer.
Questions & Comments

Lewis Jordan
Lewis.Jordan@Equifax.com