Correlation, Causality and Variable Importance

Lyle Ungar
CIS Department
University of Pennsylvania
Why do people build models?

- **ML: prediction**
  - the y-hat culture \( y = f(x; w) \)

- **Statistics: hypothesis testing (insight)**
  - the beta-hat culture \( y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \epsilon \)

- **The real world: picking the best actions**
  - Reinforcement learning
  - Causal modeling
Variable explanation/importance

- **Interpretation**
  - Find items closest to the cluster center
  - Find words closest to a vector embedding

- **Method specific or agnostic variable importance**

- **argmax_x f(x) for hidden nodes or outputs**
  - Which input (Image, document …) maximizes the p(Y=y)?
\[ \text{argmax}_x f(x) \]

(a) Extraverted.  
(b) Conscientious.

Figure 1: Example Twitter profile pictures for users scoring high in a personality trait.
Variable Importance: Regression

- \( y = 1000 \, x_1 + x_2 \)
- **Which is more important:** \( x_1 \) or \( x_2 \)?
- **How should you measure importance?**
Variable Importance: Regression

- **Univariate and multivariate are different**
  - Since features are usually highly redundant

- **True model:** \( y = x_1 + x_5 \)

- **Fit:** \( y = c_1 x_1 + c_2 x_2 + c_3 x_3 + c_4 x_4 + c_5 x_5 \)
  - with \( x_1 = x_2 = x_3 = x_4 \)

- **Giving:** \( y = \frac{1}{4} x_1 + \frac{1}{4} x_2 + \frac{1}{4} x_3 + \frac{1}{4} x_4 + c_5 x_5 \)

- **How important is** \( x_1 \)?
  - \( \frac{1}{4} \) or 1?
Generic Variable Importance

- The accuracy loss from leaving out a variable when building a model
  - What is the importance of $x_1$ in
    \[ y = c_1 x_1 + c_2 x_2 + c_3 x_3 + c_4 x_4 + c_5 x_5 \]
    with $x_1 = x_2 = x_3 = x_4$

- The accuracy loss from pegging a variable to its average value in a trained model
Random Forest Variable Importance

- Find test set error, $\text{Err}$
- Permute a variable $x_j$, find new test set error, $\text{Err}_t$
- Variable importance is the difference, $(\text{Err} - \text{Err}_t)$ divided by the standard error

From the R package for Random forests
For interpretation

- Just find correlation of each feature $x_j$ with $y$
  - But beware on nonlinear relations
  - Pearson vs. Spearman correlations

Pearson = −0.799, Spearman = −1

https://support.minitab.com/
Female words

- Mom
- Beautiful
- Love you
- Loving
- Cute
- So much
- Excited
- Shopping
- Birthday
- Yummy
- Hubby
- Super excited
-手続き
- My hair
- My heart
- AUNT
- Nail
- Bed
- Boyfriend
- Dinner
- Cookies
- Dinner
- Best friends
- Chocolate
- Birthday
- Mommy
- Christmas
- Drama
- Christmas
- Make me
- Sisters
- Cleaning
- Cleaning
- Adorable
- Cleaning
- Baking
- Hey
- Family
- Ausable
- With my
- Evening
- New
- Love them
- Baking
- Go away
- Laugh
- With my
- Daddy
- Go away
Well adjusted (anti-neurotic) words
Model Interpretation

◆ **Global**: What does this model do?
  - Requires simple models (regression, decision tree)

◆ **Local**: Why did you make this decision?
  - Decision tree: path taken
  - Regression: largest values of $w_j(x_j - \mu_j)$
  - **LIME** (Local Interpretable Model-Agnostic Explanations)
LIME algorithm

- Sample instances near the target instance
- Predict labels using full model
- Fit a sparse locally weighted regression
- The dashed line is the “explanation”
**LIME**

- Do local perturbations to $x$

![Original Image](image1.png) ![Interpretable Components](image2.png)

- Fit locally weighted model

---

**Why Should I Trust You?**: Explaining the Predictions of Any Classifier
Ribeiro Singh & Guestrin

Generate a data set of perturbed instances by turning some of the interpretable components “off” (gray).
LIME explains alternate predictions

P( ) = 0.54

P( ) = 0.07

P( ) = 0.05

- Works on SVMs, Random forests, Nnets..
- Works on text or images

Prediction probabilities

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>atheism</td>
<td>0.58</td>
<td></td>
</tr>
<tr>
<td>christian</td>
<td>0.42</td>
<td></td>
</tr>
</tbody>
</table>

Text with highlighted words

From: johnchad@triton.unm.edu (jchadwie)
Subject: Another request for Darwin Fish
Organization: University of New Mexico, Albuquerque
Lines: 11
NNTP-Posting-Host: triton.unm.edu

Hello Gang,

There have been some notes recently asking where to obtain the DARWIN fish.
This is the same question I have and I have not seen an answer on the net. If anyone has a contact please post on the net or email me.
Correlation and Causality

I used to think correlation implied causation.

Then I took a statistics class. Now I don't.

Sounds like the class helped.

Well, maybe.

https://xkcd.com/552/
High correlation between...

- Radio ownership and population in insane asylums
  - England, 20th century

- Daily ice cream consumption and rape incidents
  - US, 21st century

- Stork population and babies born
  - Germany, 20th century
New evidence for the theory of the stork.

- Höfer T, Przyrembel H, Verleger S.

Data from Berlin (Germany) show a significant correlation between the increase in the stork population around the city and the increase in [baby] deliveries outside city hospitals (out-of-hospital deliveries). However, there is no correlation between deliveries in hospital buildings (clinical deliveries) and the stork population. The decline in the number of pairs of storks in the German state of Lower Saxony between 1970 and 1985 correlated with the decrease of deliveries in that area.
Causality and Regression

◆ $y = c_1 x_1 + c_2 x_2$
  - $y$: crop yield
  - $x_1$: temperature
  - $x_2$: rainfall

Do higher temperatures cause higher crop yields?

◆ Increased temperature decreases yield?
  - $y = -0.1 x_1$

◆ Increased temperature increases yield?
  - $y = 0.2 x_1 + 0.4 x_2$
**Causality and feature selection**

\[ y = c_1 x_1 + c_2 x_2 + c_3 x_3 \]

- \( y \): customer life time value
- \( x_1 \): customer car value
- \( x_2 \): customer house value
- \( x_3 \): customer mortgage payment

**Stepwise regression selects only** \( x_3 \)

- What does this mean?
- Is this a problem?
Feedback complicates causality

- Room temperature as a function of whether the heat is on

```
Room Temp

Heater

on  off

high  low
```
Causality Matters

- If treatment is different based on gender
  - Probability of recovery based on seeing that someone is treated is different than the probability if one knows the gender.

- Causality is usually impossible to infer
  - Does treatment cause high blood pressure or high blood pressure cause treatment?
Questions

- My decision tree indicates that zip code, house price, having aluminum siding, and owning a boat are all useful predictors of purchase of projection TVs, while income is not.
  - What might be going on?
  - Is this a problem?

- Predicted polymer quality does not depend on the temperature of the reactor
  - Does this mean temperature doesn’t affect quality?
Rich Caruana found the among patients with pneumonia, those with asthma had a lower chance of dying

- What might be going on?
- Is this a problem?
Can add causality to models

Can add decision nodes to Belief Network

- Markov Process $\rightarrow$ Markov Decision Process (MDP)
- HMM $\rightarrow$ Partially Observable Markov Decision Process (POMDP)
- Neural Network $\rightarrow$ Neural Network
Belief Nets can model causality
Where is the “experimentation” in RL?
Take-Aways

◆ Machine learning finds correlation – not causality
  ● *Finding causality requires experiments*
    ■ Or talking to experts
  ● But correlations suggest possible causality
  ● Interpretable models help identify dubious causality

◆ **Actions can be added to most of our models**
  ● One then needs to both *learn the model* and *select the optimal action*
  ● Exploration in RL is experimentation