How can you teach Logistic Regression and/or Deep Learning to a non-tech 30 year old in 1 hour using online tools

Group work at Stanford is important. And hard...
Group work is central to learning
Group work is central to workforce
Peers are **human resource** that scales with class size.
How can we best work in groups? At Stanford and online

How can classrooms use peers for learning?
Learning Goals

- Know the best practices for team work
- Know the open problems in: group work
- How can AI support peer grading + peer teaching online.
Part 1: Students working in groups
What are the skills we should teach?
NASA Exercise: Survival on the Moon

Key points:
• Your spaceship crash landed on the moon
• Make a 200 mile trip to base
• Rank 15 items to take on your trip

Individually (5 mins). In a group of 4 (8 mins)
Best Practices

• Avoid conflating **identity** with ideas
• **Listen** to each other, finding the best ideas within a group is hard.
• **Nurture** the relationship
• Set clear expectations and **intentions**
• **Reflect** on your dynamics and improve!

There is always a better way.
Part 2: Online teams
Online Groupwork

- 150,000 students
- Collaborative, project based
- online
TeamRank

Team Awesome

None of the other teams seemed to fit my needs, so I created one of my own. If you need someone (or a few someones) who can do it well, a lot of the teams look to be school specific and I am here on my...

More Info
Team formation characteristics:
Teamwork Goals

Team Members:
- Jessica Leibovitz, Englewood, CO, United States
- Angela Shortall, Marietta, GA, United States
- Veer Singh, Glenville, MD, United States
- Marissa Peled, Westfield, CO, United States
- Amy Fazakara Syed Ahmad, United Arab Emirates

Recent Activity
- I am getting very frustrated! I just spent an hour putting information in for assignment 1, and for some reason, it just closed on me and lost all of the information I spent in. This is the second time this has happened, and it's making me want to just shut this class. Has this happened to anyone else, and does anyone have any suggestions on why this is happening?

Survival Curves
Team Influence in NovoEd

Survival Curves
Mean activity & Leader dropout
Mean activity & Teammate dropout

Graph showing survival curves with participation week on the x-axis and percentage punks on the y-axis.
NoVoEd Study: Group work matters?

Figure 1. Graduation Rate by Cohort

Figure 2. Engagement by Cohort

Context and research methodology

This research was conducted in 2014 by Chuck Eesley, Assistant Professor of Management Science & Engineering at Stanford University. The data is from 26,248 students in Technology Entrepreneurship, an eight-week free course. The analysis utilizes a multivariate regression format, with dependent variables of various engagement and satisfaction measures, independent variables including collaboration type, and control variables for demographics, engagement level, and more.

Born into Chaos: the Role of the Founding Environment and the Performance Impacts of Founding Team Composition

Charles Eesley, D. Carrington Molley and Wesley Wu-Yi Koo

Published Online: 28 Jan 2019

Abstract

How does a firm's founding environment shape its founding team's ability to drive performance? Upper echelon research predicts that in a volatile industry environment having a diverse top management team is beneficial. In contrast, we argue that a diverse team in the presence of environmental volatility at founding might hamper a team's performance in the long run. Using a sample of 1,383 firms, we find that a diverse founding team in a volatile founding environment significantly lowers the chance of favorable exit for firms. That relationship is particularly pronounced for firms whose environment stabilized over time and for firms with quick product cycles. Our results suggest that founders need to be aware of the long-term impacts of matching team composition to the founding environment.
Part 3: Peers as teachers
Tuned Peer Grading

63,000+ peer grades
<table>
<thead>
<tr>
<th>Grader</th>
<th>Gradee</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>user012</td>
<td>user734</td>
<td>72</td>
</tr>
<tr>
<td>user012</td>
<td>user384</td>
<td>99</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>user384</td>
<td>user001</td>
<td>5</td>
</tr>
</tbody>
</table>

63,000+ peer grades
Peer Graders Aren’t Very Accurate

Some students were getting very erroneous grades

Error is based on ground truth assignments. Results are across all students (~10,000 submissions)
Aside: Research pre-neural nets

Make a model of the world

Infer unobserved variables given data. MCMC or EM

Analysis
Some students were getting very erroneous grades.

Error is based on ground truth assignments. Results are across all students (~10,000 submissions).

Probabilistic Model

\[
\begin{align*}
(\text{Reliability}) & \quad \tau_v \sim \mathcal{G}(\alpha_0, \beta_0) \text{ for every grader } v, \\
(\text{Bias}) & \quad b_v \sim \mathcal{N}(0, 1/\eta_0) \text{ for every grader } v, \\
(\text{True score}) & \quad s_u \sim \mathcal{N}(\mu_0, 1/\gamma_0) \text{ for every user } u, \text{ and} \\
(\text{Observed score}) & \quad z_u^v \sim \mathcal{N}(s_u + b_v, 1/\tau_v), \\
& \quad \text{for every observed peer grade.}
\end{align*}
\]

Novel Model:

\[
z_u^v \sim \mathcal{N}\left(s_u + b_v, \frac{1}{\theta_1 s_u + \theta_0}\right)
\]
Some students were getting very erroneous grades.

Error is based on ground truth assignments. Results are across all students (~10,000 submissions).
Some students were getting very erroneous grades.

Error is based on ground truth assignments. Results are across all students (~10,000 submissions).

99% within 10pp
Markov Chain

MCMC

Monte Carlo
<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[1, 1, 101.0, 1]</td>
<td>[1, 1, 101.0, 1]</td>
<td>[0, 1, 101.0, 0]</td>
<td>[0, 0, 101.0, 0]</td>
<td>[1, 0, 101.0, 1]</td>
</tr>
<tr>
<td>[1, 0, 101.0, 0]</td>
<td>[1, 0, 101.0, 0]</td>
<td>[1, 0, 101.0, 1]</td>
<td>[1, 0, 101.0, 1]</td>
<td>[1, 0, 101.0, 1]</td>
</tr>
<tr>
<td>[1, 1, 101.0, 1]</td>
<td>[1, 1, 101.0, 1]</td>
<td>[1, 1, 101.0, 1]</td>
<td>[1, 1, 101.0, 1]</td>
<td>[1, 1, 101.0, 1]</td>
</tr>
<tr>
<td>[1, 1, 101.0, 1]</td>
<td>[1, 1, 101.0, 1]</td>
<td>[1, 1, 101.0, 1]</td>
<td>[1, 1, 101.0, 1]</td>
<td>[1, 1, 101.0, 1]</td>
</tr>
<tr>
<td>Pr(Flu) = 0.9773</td>
<td>Pr(Flu) = 0.9773</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MCMC is a way to sample from the “joint” with conditioned variables fixed.

Each one of these is one posterior sample:

Which only requires you to derive the probability of each variable given assignments to the rest.
Aside: MCMC Derivation

**Chris Piech, Jonathan Huang, Zhenghao Chen, Chuong Do, Andrew Ng, Daphne Koller**

**Derivation of updates.** We examine the problems of sampling $s_u$ and $\tau_v$ separately. Consider now a fixed user $u_i$. We derive the sampling step for $s_u$ as follows:

$$s \sim P(s_{u_i}|MB(s_{u_i})), \quad \propto P(s_{u_i}|\mu_0, \gamma_0) \cdot \prod_{v:v \rightarrow u_i} P(z_{u_i}^v|s_{u_i}, b_v, \tau_v), \quad \propto \exp\left(-\frac{1}{2}\gamma_0(s_{u_i} - \mu_0)^2 + \sum_{v:v \rightarrow u_i} \left(-\frac{1}{2}\tau_v \left(z_{u_i}^v - (s_{u_i} + b_v)\right)^2\right)\right), \quad \propto \exp\left(-\frac{1}{2} \left[\gamma_0(s_{u_i} - \mu_0)^2 + \sum_{v:v \rightarrow u_i} \tau_v \left(z_{u_i}^v - (s_{u_i} + b_v)\right)^2\right]\right).$$

---

**Anatomy of a beautiful equation**

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- $\mathcal{N}(\mu, \sigma^2)$
- Probability density at $x$
- A constant
- "Exponential"
- The distance to the mean
- Sigma shows up twice
- a constant
- $f(x) = K \cdot \exp\left[-\frac{1}{2} \tau (x - \mu)\right]$
\[ \propto \exp \left( -\frac{1}{2} \left[ \gamma_0 (s_{u_i} - \mu_0)^2 + \sum_{v:v \rightarrow u_i} \tau_v (z^v_{u_i} - (s_{u_i} + b_v))^2 \right] \right). \]

Recall

\[ f(x) = K \cdot \exp \left( -\frac{1}{2} \tau (x - \mu) \right) \]

The expression inside the exponent is quadratic — we thus complete the square, obtaining:

\[
\gamma_0 (s_{u_i} - \mu_0)^2 + \sum_{v:v \rightarrow u_i} \tau_v (z^v_{u_i} - (s_{u_i} + b_v))^2
\]

\[
= \text{const.} + \gamma_0 (s_{u_i}^2 - 2\mu_0 s_{u_i}) + \sum_{v:v \rightarrow u_i} \tau_v ((s_{u_i} + b_v)^2 - 2z^v_{u_i} (s_{u_i} + b_v)),
\]

\[
= \text{const.} + \left( \gamma_0 + \sum_{v:v \rightarrow u_i} \tau_v \right) s_{u_i}^2 - 2 \left( \gamma_0 \mu_0 + \sum_{v:v \rightarrow u_i} \tau_v (z^v_{u_i} - b_v) \right) s_{u_i},
\]

\[
= \text{const.} + R \left( s_{u_i} - \frac{1}{R} \left( \gamma_0 \mu_0 + \sum_{v:v \rightarrow u_i} \tau_v (z^v_{u_i} - b_v) \right) \right)^2,
\]

(where \( R = \gamma_0 + \sum_{v:v \rightarrow u_i} \tau_v \)).

Complete the Square:

... take a Quadratic Equation like this:

\[ ax^2 + bx + c = 0 \]

and turn it into this:

\[ a(x+d)^2 + e = 0 \]
Utility beyond “more accurate grades?”
Grading Sweet Spot

"sweet spot of grading":
~ 20 minutes

Grade Time Vs Accuracy

Chris Piech, Jon Huang, Daphne Koller. Tuned Models of Peer Assessments. EDM 2013
Unsolved: students don’t want to peer grade
Unsolved: Just the start of group work online...