Unstructured Data Analysis

Lecture 11: Wrap-up predictive model evaluation, classical classifiers; intro to neural nets & deep learning

George Chen
Today

• Wrap-up coverage of how to evaluate whether a predictive model is good & classical classifiers

• In many datasets (especially small, structured ones), neural nets & deep learning could work poorly… in such cases, often decision-tree-based methods can work well

  \textit{random forests, gradient boosting (e.g., XGBoost)}

• Start coverage of neural nets & deep learning
Decision Trees & Forests
Example Made-Up Data

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Weight (lb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>100</td>
</tr>
<tr>
<td>20</td>
<td>200</td>
</tr>
<tr>
<td>20</td>
<td>300</td>
</tr>
<tr>
<td>40</td>
<td>100</td>
</tr>
<tr>
<td>40</td>
<td>200</td>
</tr>
<tr>
<td>40</td>
<td>300</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>50</td>
<td>200</td>
</tr>
<tr>
<td>50</td>
<td>300</td>
</tr>
</tbody>
</table>

Red: diabetic
Blue: not diabetic
Learning a Decision Tree

- Many ways: general approach actually looks a lot like divisive clustering but accounts for label information

- I’ll show one way (that nobody actually uses in practice) but it’s easy to explain
Learning a Decision Tree

1. Pick a random feature (either age or weight)

2. Find threshold for which red and blue are as "separate as possible" (on one side, mostly red; on other side, mostly blue)
Learning a Decision Tree

Within each side, recurse until a termination criterion is reached!

Example termination criteria: \( \geq 90\% \) points within region has same label, number of points within region is < 5

Note: label within each region is majority vote

Red: diabetic
Blue: not diabetic
For a new person with feature vector (age, weight), easy to predict!
Decision Forest for Classification

- Typically, a decision tree is learned with randomness (e.g., we randomly chose which feature to threshold)
  - by re-running the same learning procedure, we can get different decision trees that make different predictions!
- For a more stable prediction, use many decision trees

Final prediction: majority vote of the different trees’ predictions

Learn each tree separately using same training data

New test data point

Tree 1
- diabetic

Tree 2
- not diabetic

Tree 3
- diabetic

... Tree \( T \)
- diabetic
Decision Forest for Classification

Randomly sample (with replacement) $n$ points

New test data point

Tree 1

diabetic

Tree 2

not diabetic

Tree 3

diabetic

…

Tree $T$

diabetic

Question: What happens if all the trees are the same?

Adding randomness can make trees more different!

• **Random Forest:** randomize training data used for each tree, randomly choose a few features to try to split on (and among these features, choose the best one to split on)
Back to the demo
Another Way to Benchmark

• In the demo: we just saw that we can compare test set prediction accuracy across different algorithms and also look at confusion matrices

• For binary classification, we can do a more detailed analysis
Binary Classification: ROC Curves

For simplicity, think of the random forest for now

New test data point

Tree 1

positive

Tree 2

negative

Tree 3

positive

Tree $T$

positive

Final prediction: majority vote of the different trees’ predictions

$\geq 50\%$ of trees need to say positive for final prediction to be positive

We can vary this 50% threshold!
Binary Classification: ROC Curves

Error rates are computed on test data.
Binary Classification: ROC Curves

A classifier with the green curve is better than the one with the blue curve.
Binary Classification: ROC Curves

It’s possible that algorithms are better in different regimes.
Binary Classification: ROC Curves

The ideal curve (typically impossible to achieve)

Area under the curve (AUC) is a popular metric for comparing algorithms (higher is better)
Binary Classification: ROC Curves

What we just saw:

- For a classifier that we can set the threshold probability to different values, we can plot an ROC curve
- True positive rate (TPR) and false positive rate (FPR) are evaluated on test data

Other variants are possible:

- Plot precision vs recall instead of TPR vs FPR
- Can actually plot ROC/precision-recall curves sweeping over hyperparameters aside from threshold probability!
- For ROC/precision-recall, rather than evaluating on test data, can evaluate on validation data during training to help choose hyperparameters
Binary Classification: ROC Curves

Different hyperparameter settings (need not be only for threshold probability)

Can also be computed on validation data instead of test data!
Intro to Neural Nets & Deep Learning
Over 10 million images, 1000 object classes

2011: Traditional computer vision achieves accuracy ~74%
2012: Initial deep neural network approach accuracy ~84%
2015 onwards: Deep learning achieves accuracy 96%+

Deep Learning

Extremely useful in practice:

• Near human level image classification (including handwritten digit recognition)

• Near human level speech recognition

• Improvements in machine translation, text-to-speech

• Self-driving cars

• Better than humans at playing Go
Google DeepMind’s AlphaGo vs Lee Sedol, 2016
DeepMind’s StarCraft 2 AI is now better than 99.8 percent of all human players

*AlphaStar is now grandmaster level in the real-time strategy game*

By Nick Statt | @nickstatt | Oct 30, 2019, 2:00pm EDT
Turing Award Won by 3
Pioneers in Artificial Intelligence

From left, Yann LeCun, Geoffrey Hinton and Yoshua Bengio. The researchers worked on key developments for neural networks, which are reshaping how computer systems are built. From left, Facebook, via Associated Press; Aaron Vincent Elkaim for The New York Times; Chad Buchanan/Getty Images

By Cade Metz

March 27, 2019
Is it all hype?
panda  \( \sim 58\% \) confidence

adversarial noise

\[ + 0.007 \times \]

gibbon  \( \sim 99\% \) confidence

Fooling Neural Networks in the Physical World with 3D Adversarial Objects

31 Oct 2017 · 3 min read — shared on Hacker News, Lobsters, Reddit, Twitter

We’ve developed an approach to generate 3D adversarial objects that reliably fool neural networks in the real world, no matter how the objects are looked at.

Neural network based classifiers reach near-human performance in many tasks, and they’re used in high risk, real world systems. Yet, these same neural networks are particularly vulnerable to adversarial examples, carefully perturbed inputs that cause
a cat is sitting on a toilet in a bathroom

Source: Gizmodo article “This Neural Network's Hilariously Bad Image Descriptions Are Still Advanced AI”. September 16, 2015. (They’re using the NeuralTalk image-to-caption software.)
<table>
<thead>
<tr>
<th>General</th>
<th>VIEW DOCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>no person</td>
<td>0.991</td>
</tr>
<tr>
<td>beach</td>
<td>0.990</td>
</tr>
<tr>
<td>water</td>
<td>0.985</td>
</tr>
<tr>
<td>sand</td>
<td>0.981</td>
</tr>
<tr>
<td>sea</td>
<td>0.980</td>
</tr>
<tr>
<td>travel</td>
<td>0.978</td>
</tr>
<tr>
<td>seashore</td>
<td>0.972</td>
</tr>
<tr>
<td>summer</td>
<td>0.954</td>
</tr>
<tr>
<td>sky</td>
<td>0.946</td>
</tr>
<tr>
<td>outdoors</td>
<td>0.944</td>
</tr>
<tr>
<td>ocean</td>
<td>0.936</td>
</tr>
</tbody>
</table>

*Source: Pietro Perona*

**cow is not among top objects found!**
<table>
<thead>
<tr>
<th>General Concept</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>group</td>
<td>0.979</td>
</tr>
<tr>
<td>adult</td>
<td>0.977</td>
</tr>
<tr>
<td>people</td>
<td>0.976</td>
</tr>
<tr>
<td>furniture</td>
<td>0.960</td>
</tr>
<tr>
<td>room</td>
<td>0.957</td>
</tr>
<tr>
<td>business</td>
<td>0.903</td>
</tr>
<tr>
<td>indoors</td>
<td>0.901</td>
</tr>
<tr>
<td>man</td>
<td>0.896</td>
</tr>
<tr>
<td>seat</td>
<td>0.895</td>
</tr>
</tbody>
</table>

elephant is not among top objects found!

Source: David Lopez-Paz
Another AI Winter?

~1970’s: First AI winter over symbolic AI

~1980’s: Second AI winter over “expert systems”

Every time: Lots of hype, explosion in funding, then bubble bursts
Artificial Intelligence—The Revolution Hasn’t Happened Yet

Artificial Intelligence (AI) is the mantra of the current era. The phrase is intoned by technologists, academicians, journalists and venture capitalists.

What is deep learning?