Microsoft

MLADS

MACHINE LEARNING, AI,
AND DATA SCIENCE CONFERENCE

JUNE, 2020

Microsoft

Generalization, Utility, and Experimentation: ML Concepts for Making Better Business Decisions: Section 2: Automating Decisions

Robert Horton
John-Mark Agosta
Mario Inchiosa

Automating Decisions: Quantifying quality/quantity tradeoffs in ML

Robert Horton

Session goals

1. **Learn how machine learning (ML) differs** from traditional software engineering
2. See how ML fits in the context of **making better business decisions**
3. Understand why causal relationships matter in data analysis, and **why we still need to do experiments**

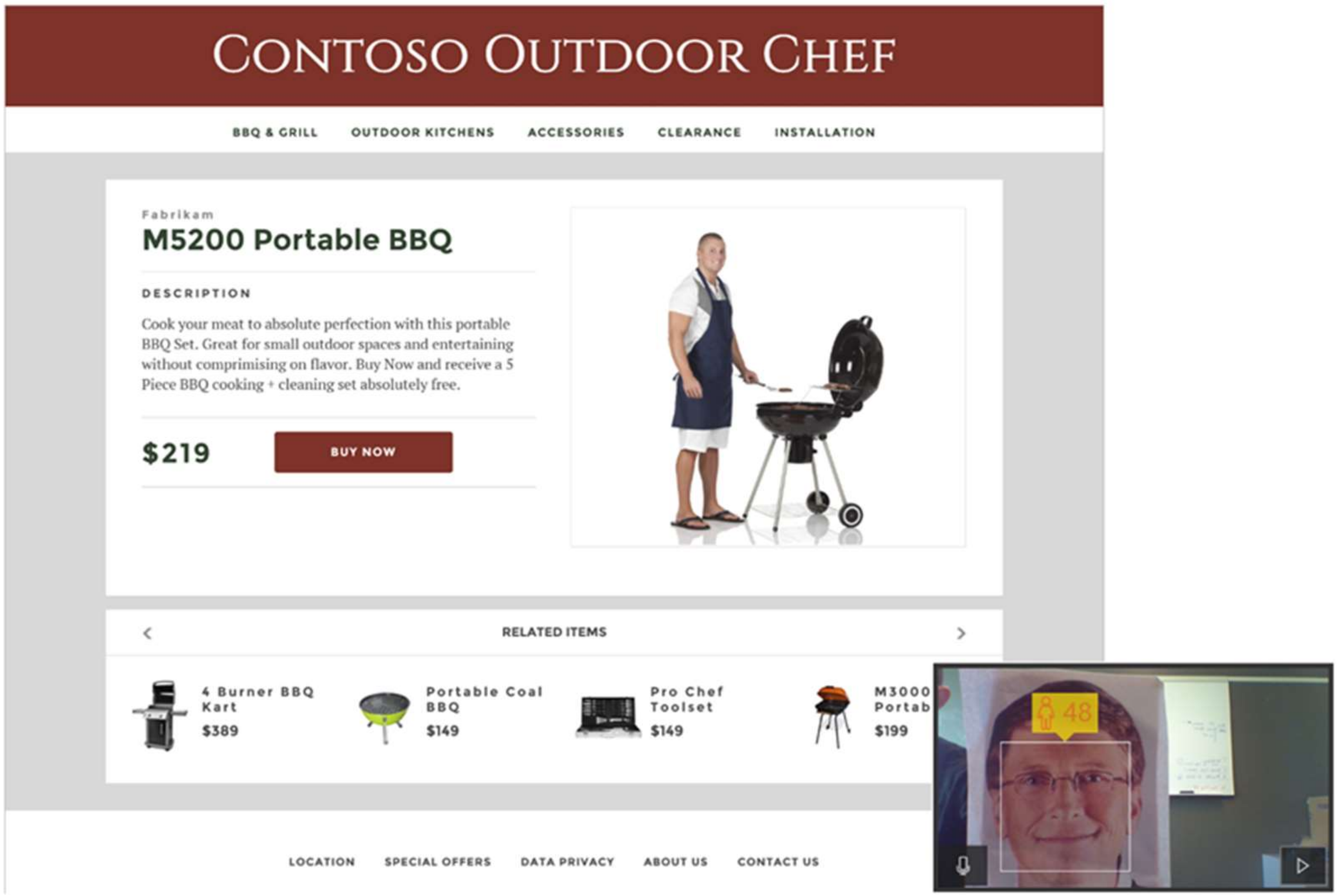
Using classifiers in business decisions

- You already make mistakes
- ML models make mistakes too
  - Can ML help you make fewer or less expensive mistakes?

- Economic Utility of a Binary Classifier
  - How much do we gain by correctly identifying positive and negative cases?
  - How much does each type of mistake cost us?


How-Old.net
HOW OLD DO I LOOK? #HowOldRobot
41
Stay tuned for the Windows Central Podcast!

Mall Kiosk


CONTOSO OUTDOOR CHEF
BBQ & GRILL
OUTDOOR KITCHENS
ACCESSORIES
CLEARANCE
INSTALLATION
Fabrikam
M5200 Portable BBQ
DESCRIPTION
Cook your meat to absolute perfection with this portable BBQ Set. Great for small outdoor spaces and entertaining without comprimising on flavor. Buy Now and receive a 5 Piece BBQ cooking + cleaning set absolutely free.
$219
BUY NOW
RELATED ITEMS
4 Burner BBQ Kart
$389
Portable Coal BBQ
$149
Pro Chef Toolset
$149
M3000 Portab
$199
48
LOCATION
SPECIAL OFFERS
DATA PRIVACY
ABOUT US
CONTACT US

Confusion Matrix

observed predicted	True TPR	False FPR
True	TP	FP
False	FN	TN

<table>
<tr><td colspan="2"></td><td colspan="2">True condition</td><td colspan="3"></td></tr>
<tr><td></td><td>Total population</td><td>Condition positive</td><td>Condition negative</td><td>Prevalence = $\frac{\Sigma \text{ Condition positive}}{\Sigma \text{ Total population}}$</td><td colspan="2">Accuracy (ACC) = $\frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$</td></tr>
<tr><td rowspan="2">Predicted condition</td><td>Predicted condition positive</td><td>True positive</td><td>False positive,
Type I error</td><td>Positive predictive value (PPV), Precision = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Predicted condition positive}}$</td><td colspan="2">False discovery rate (FDR) = $\frac{\Sigma \text{ False positive}}{\Sigma \text{ Predicted condition positive}}$</td></tr>
<tr><td>Predicted condition negative</td><td>False negative,
Type II error</td><td>True negative</td><td>False omission rate (FOR) = $\frac{\Sigma \text{ False negative}}{\Sigma \text{ Predicted condition negative}}$</td><td colspan="2">Negative predictive value (NPV) = $\frac{\Sigma \text{ True negative}}{\Sigma \text{ Predicted condition negative}}$</td></tr>
<tr><td colspan="2" rowspan="2"></td><td>True positive rate (TPR), Recall, Sensitivity, probability of detection, Power = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$</td><td>False positive rate (FPR), Fall-out, probability of false alarm $= \frac{\Sigma \text{ False positive}}{\Sigma \text{ Condition negative}}$</td><td>Positive likelihood ratio (LR+) = $\frac{\text{TPR}}{\text{FPR}}$</td><td rowspan="2">Diagnostic odds ratio (DOR) = $\frac{\text{LR+}}{\text{LR−}}$</td><td rowspan="2">$F_1$ score = $2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$</td></tr>
<tr><td>False negative rate (FNR), Miss rate $= \frac{\Sigma \text{ False negative}}{\Sigma \text{ Condition positive}}$</td><td>Specificity (SPC), Selectivity, True negative rate (TNR) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$</td><td>Negative likelihood ratio (LR−) = $\frac{\text{FNR}}{\text{TNR}}$</td></tr>
</table>

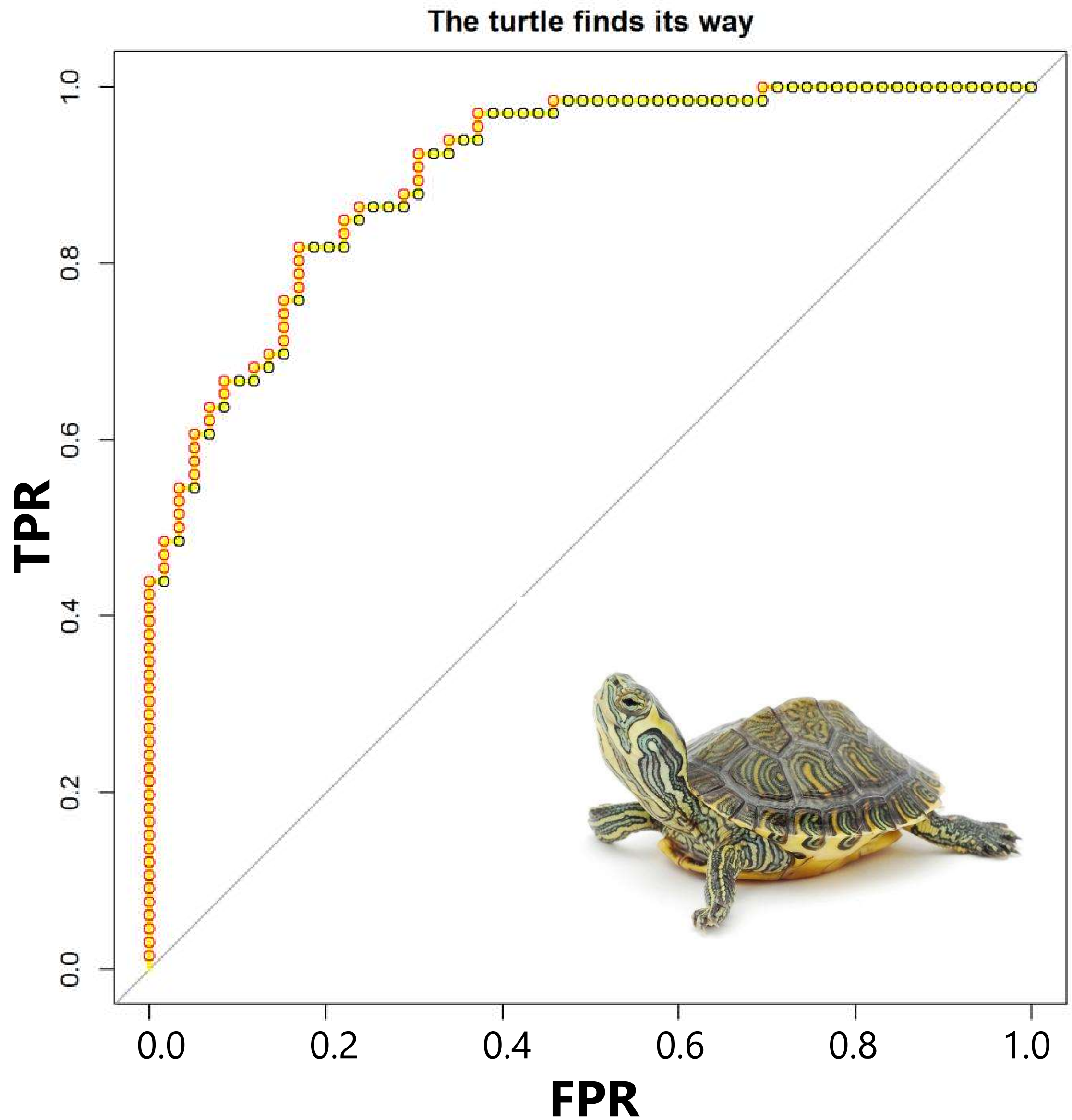
https://en.wikipedia.org/wiki/Confusion_matrix

ROC Curve:

- sort test cases by their score from the model
- march along the sequence, stepping up for positives and right for negatives

This is the same as scanning across possible cutoff threshold values.

The slope of the curve shows the concentration of positives.



Linear Utility Model

```
(TP_value * N * tpr * P) +                # sold
(FP_value * N * fpr * (1 - P)) +          # refunded
(TN_value * N * (1 - fpr) * (1 - P)) +    # trashed
(FN_value * N * (1 - tpr) * P)            # wasted
```

`N`: number of units

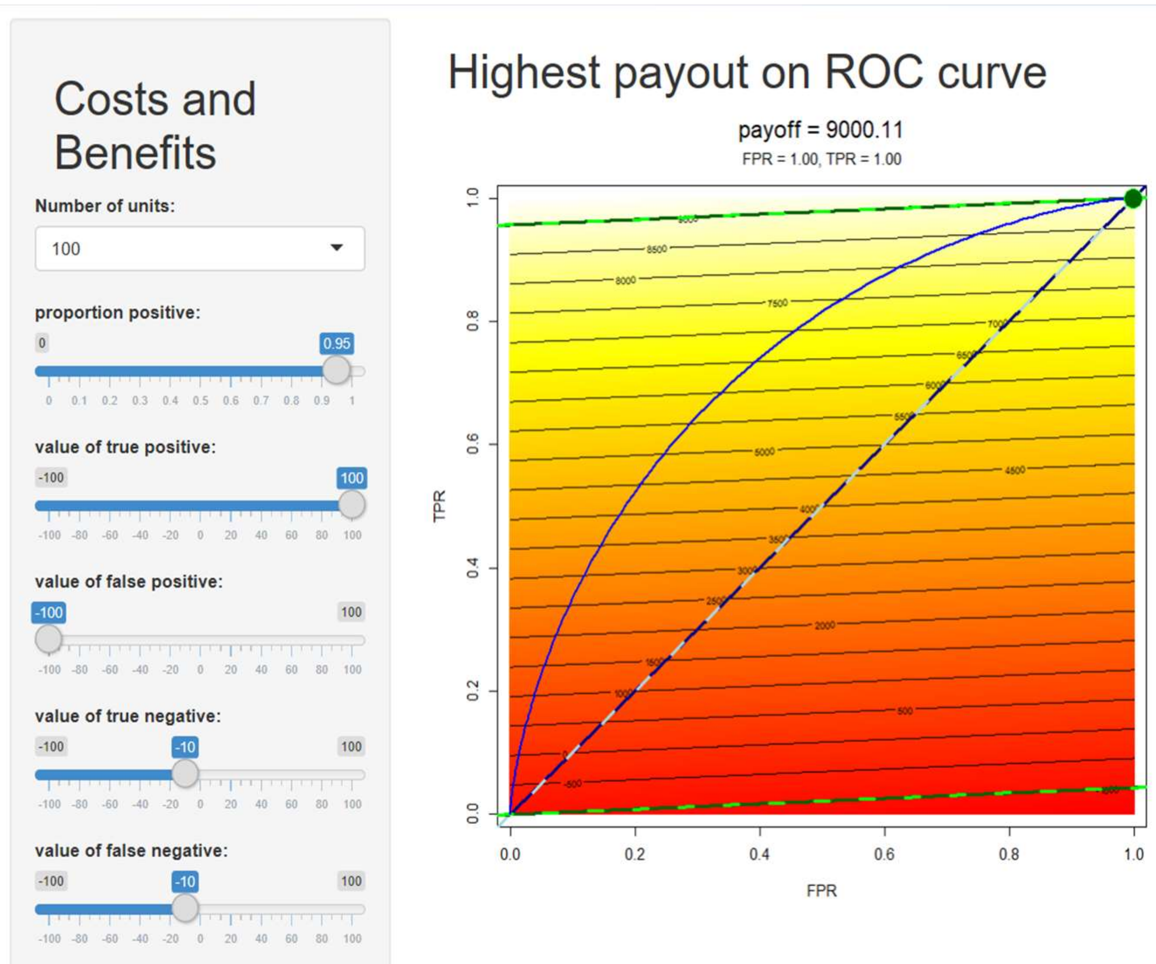
`P`: overall proportion positive

`TP_value`, `FP_value`, `TN_value`, `FN_value`: values (or costs) assigned to TP, FP, TN and FN cases

`tpr`, `fpr`: true positive and false positive rates

Economic Utility in ROC Space

The settings on the left are used to compute the “payoff” value of every point on the plane of true positive rate (TPR) vs. false positive rate (FPR); this is used to color the background. The parallel black lines are “lines of indifference”, showing contours on the cost surface (since this cost model is linear, these are parallel lines). The ROC curve is shown in blue, and the green spot shows the point on the curve with the highest payoff. In this case the highest payoff is at [1.0, 1.0], where all cases are considered positive, and the classifier is not useful.



https://ml4managers.shinyapps.io/ROC_utility/

Resources

- Github repo: https://github.com/microsoft/datascience4managers

- Shiny Apps:
  - https://ml4managers.shinyapps.io/ROC_utility/
  - https://ml4managers.shinyapps.io/effects_of_x_and_z/

- Economic Utility Functions Meet ROC Curves: Deciding on a Cutoff Threshold for Binary Classification. Siddarth Ramesh and Robert Horton, MLADS November 14, 2018. https://resnet.microsoft.com/video/4248
  - https://github.com/Azure/utility_functions_in_ROC_space
  - https://ml4managers.shinyapps.io/ROC_utility/

Thank you for attending the MLADS Conference and helping to build a strong community

To find recordings, presentations, and other resources from the event, go to: https://aka.ms/spring2020mlads

Overloaded Terms in Data Science

- **model**
  - *Statistics*: (data model) a description of a system using mathematical concepts and language (with statistical assumptions about sample generation.)
  - *ML*: (algorithmic model) data generation is a black box; the algorithm is about how to find correlations between features and outcomes.
- **inference**
  - *Statistics*: ’the process of using data analysis to deduce properties of an underlying probability distribution’ (to infer properties of the population). (Wikipedia)
  - *ML*: scoring or classifying new cases
- **experiment**
  - *Statistics*: measuring the state or value of a dependent variable when an independent variable is perturbed under controlled conditions in order to establish a cause and effect relationship.
  - *ML*: Try a bunch of algorithms, hyperparameter settings, etc. to see how they affect performance.
- **regression**
  - *English:* 'a return to a former or less developed state.'
  - *Statistics:* (regression toward the mean).
  - *ML:* prediction of a continuous-valued outcome. Contrasted with classification.