Chapter 8 Visualizing Model Performance

1/22/20 7:40 PM

Ranking instead of Classifying

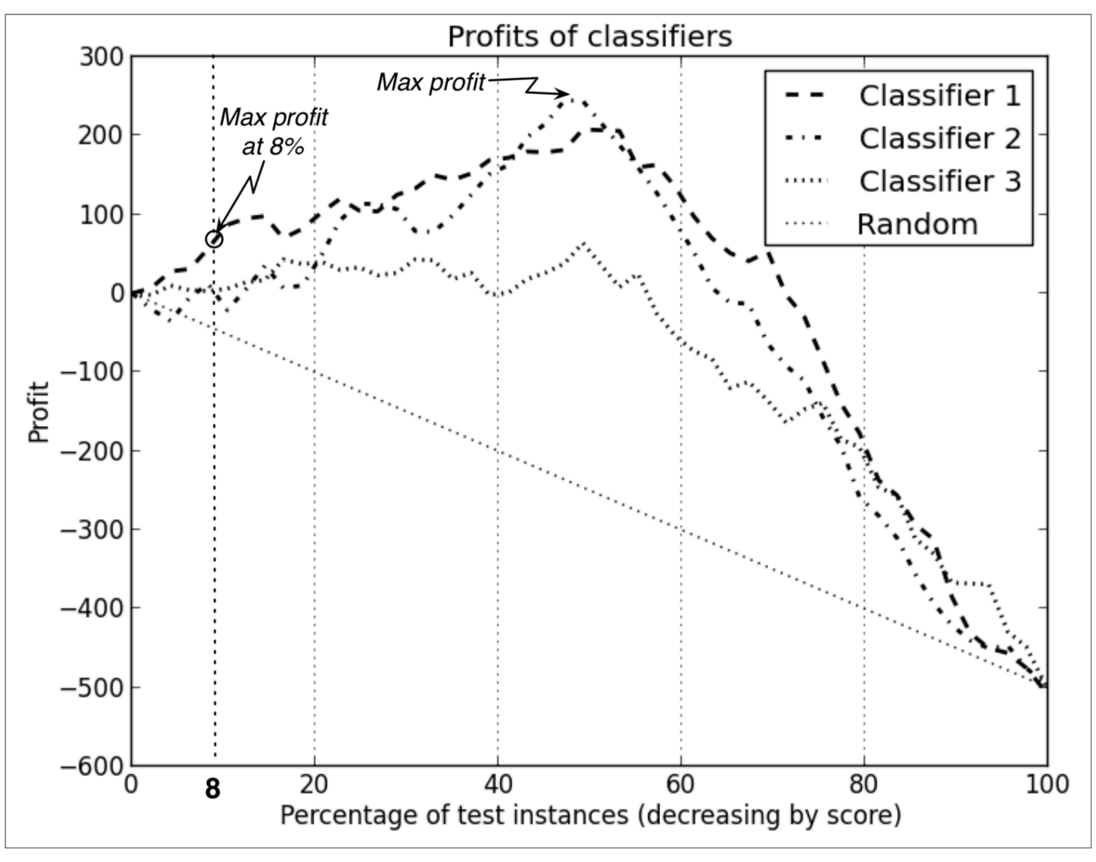
- Although we can't get accurate estimate of probability, a ranking of scores will still help us to make decision. For example, in targeted marketing when training data is not sufficient, the estimate of likelihood is not accurate but still helps.
- If we have budget for actions, and assuming cost-benefit matrix is constant, then the ranking of likelihood of an instance being positive is sufficient to make decision.

Two major questions we ask:

1. How do we compare different rankings?
2. How do we choose a proper threshold?

Profit Curve

- With a ranking classifier, we can produce a list of instances and predicted scores
- Measure the expected profit that result from each successive cut-point in the list
- Graph the expected profit against all possible cut-points



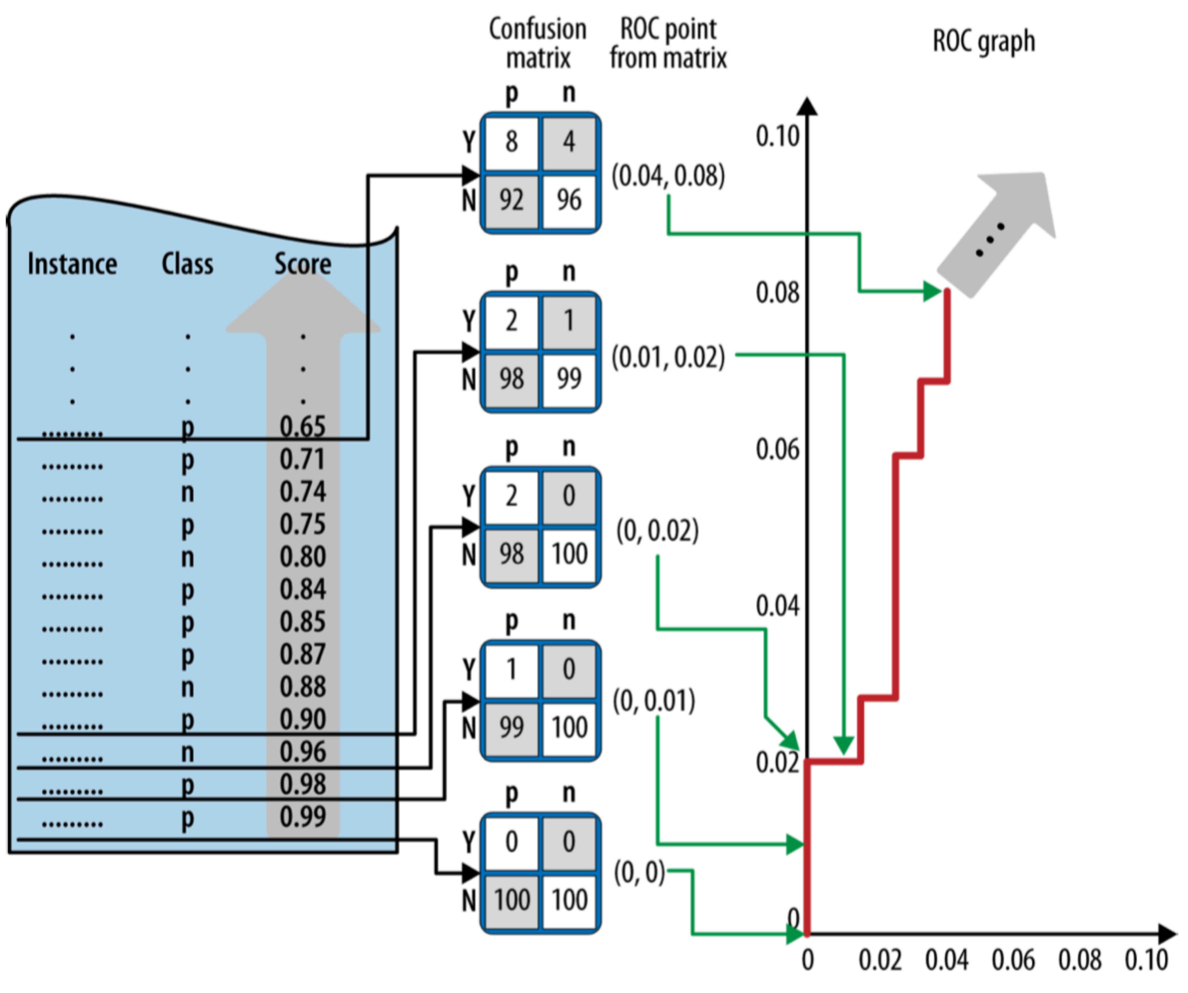
Example: if budget is $40000, each target costs $5. We have at most 8000 customers available to reach out. Given total population is 100,000, the targeting rate is 8%. Model 1 performs best among the three.

Conditions for using profit curve:

1. Class prior (base rate) should be known and expected to be stable. The expected value of profit for each instance is sensitive to base rate.
2. Costs and benefits must be available in order to calculate profit

Receiver Operating Characteristics (ROC)

- True positive rate (y axis) v.s. False positive rate (x axis)
- (0,0): all instances classified as negative, so TPR = FPR = 0
- (1,1): all instances classified as positive, so TPR = FPR = 1
- Diagonal line: random guessing, with different probability of assigning positive class
- When moving classification threshold (score), each point on ROC is generated



- **Advantage of ROC: the performance of classifier won't change from the condition, i.e. independent of class ratio and cost & benefits**

Area under ROC curve (AUC)

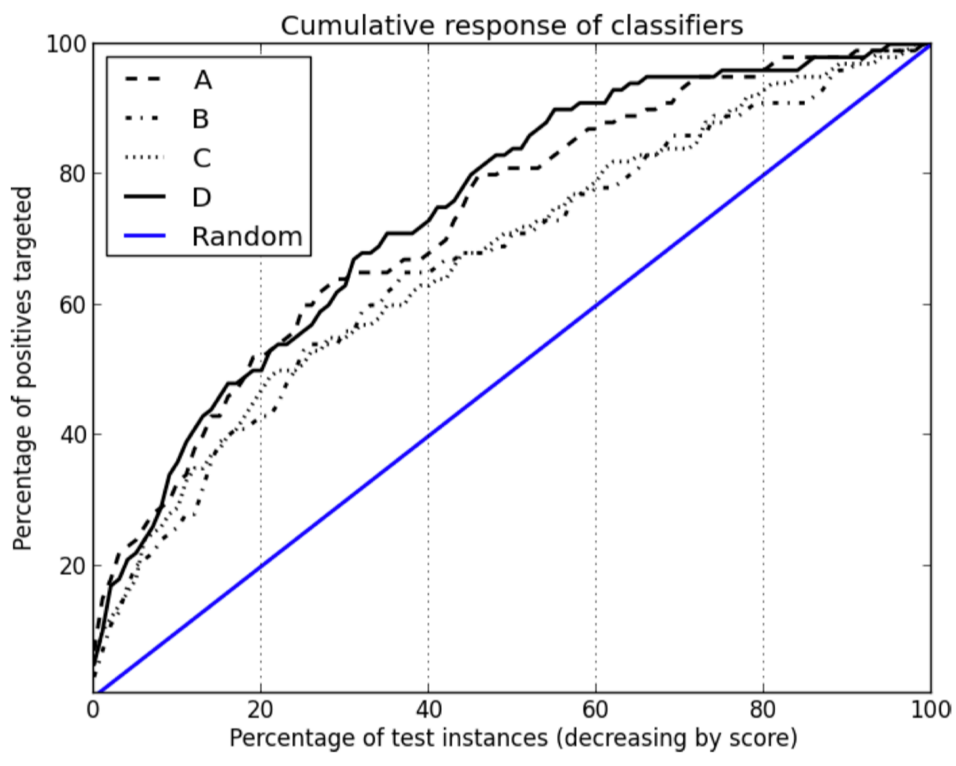
- Equivalent to: Mann-Whitney Wilcoxon measure, as well as Gini index
- Probability that a randomly chosen positive instance will be ranked ahead of a randomly chosen negative instance.

Cumulative Response and Lift Curves

Might not have all nice properties as ROC, but are more intuitive

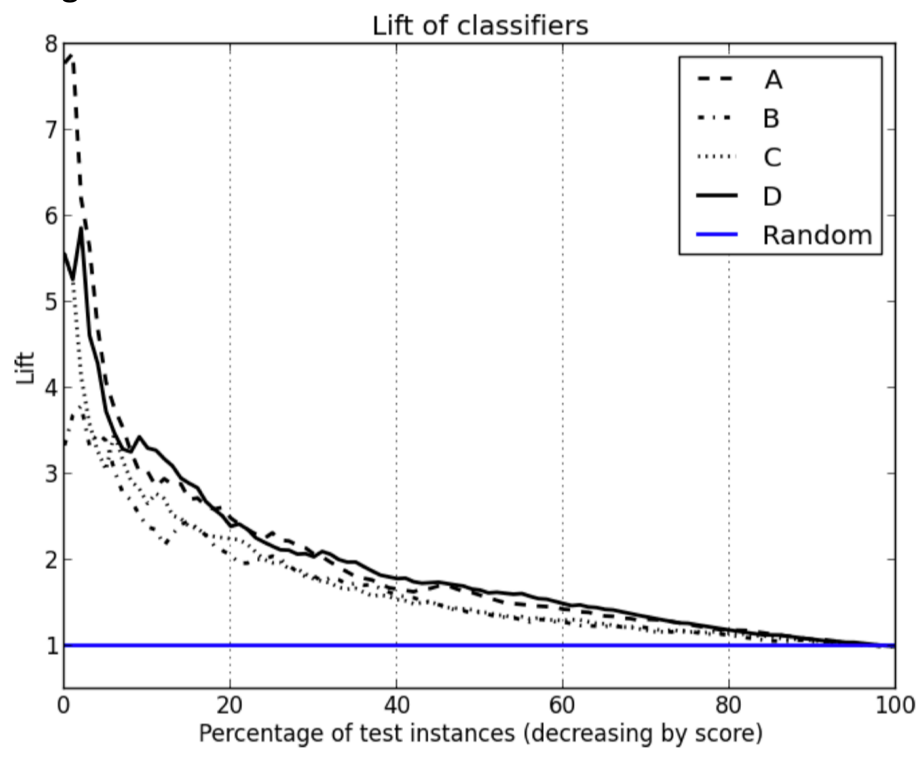
Cumulative Response Curve

- Plot true positive rate (y-axis) as function of percentage of population that is targeted (x-axis)
- Diagonal line y=x still represents random performance



Lift Curve

- It shows the advantage over random guessing
- Should be TPR minus diagonal line value from cumulative response curve
- A common claim: "Our model gives a two times (or a 2X) lift"
  - For a chosen threshold, the lift curve shows that the model's targeting is twice as good as random



** Cumulative Response Curve and Lift Curve both assumes the test data has exactly same target ratio priors as the population where model is to be applied

For example, if down-sample is applied during training, the relationships between values and axis are no longer accurate, although the shape of curve might still be informative