Profit Curves

Clayton W. Schupp

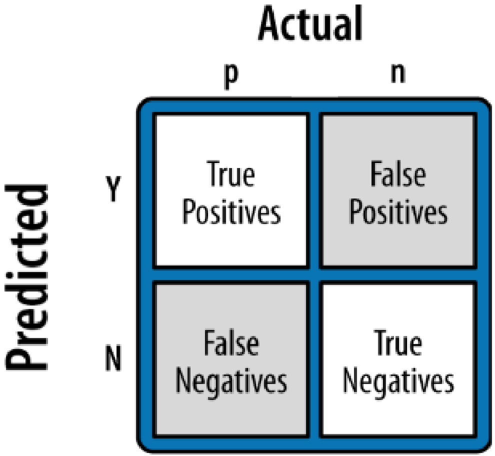
Galvanize

Cost-Benefit Information

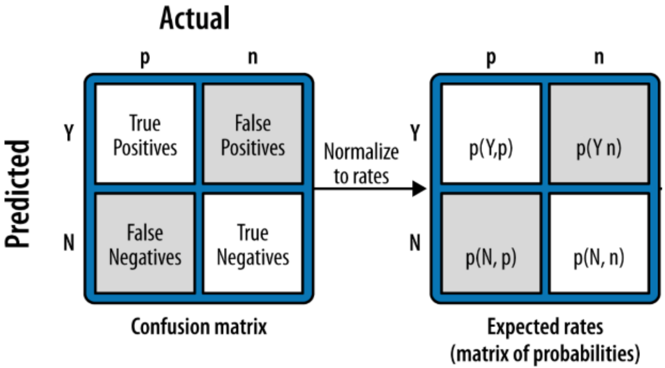
- ROC Curves alone assume an equal cost due to misclassification
- However
  - Different kinds of errors have different costs
  - Correct classifications could also have different benefits

Profit Curves allow us to compare models and select the one that will maximize profit for a specified cost-benefit

Start with the Confusion Matrix

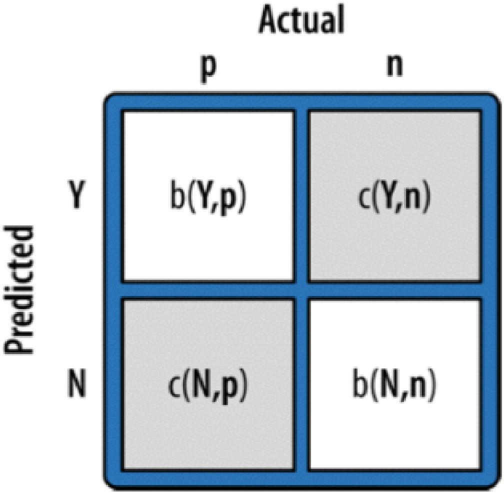


Normalize Confusion Matrix to Rates



The joint probabilities are found by taking the cell counts divided by total sample size.

Add in Cost-Benefit Matrix



Calculating Expected Profit

By combining information from the Confusion Matrix and the Cost-Benefit Matrix, we can calculate the Expected Profit:

$$
\begin{aligned}
E[Profit] \quad = \quad & P(Y,p)\cdot b(Y,p) + P(Y,n)\cdot c(Y,n) + \\
& P(N,p)\cdot c(N,p) + P(N,n)\cdot b(N,n)
\end{aligned}
$$

Expected Profit can also be calculated in terms of the true positive rate, false positive rate, true negative rate, false positive rate and the marginal positive and negative probabilites:

$$
\begin{aligned}
E[Profit] \quad = \quad & P(p)\cdot[P(Y|p)\cdot b(Y,p) + P(N|p)\cdot c(N,p)] + \\
& P(n)\cdot[P(Y|n)\cdot c(Y,n) + P(N|n)\cdot b(N,n)]
\end{aligned}
$$

Building the Profit Curve

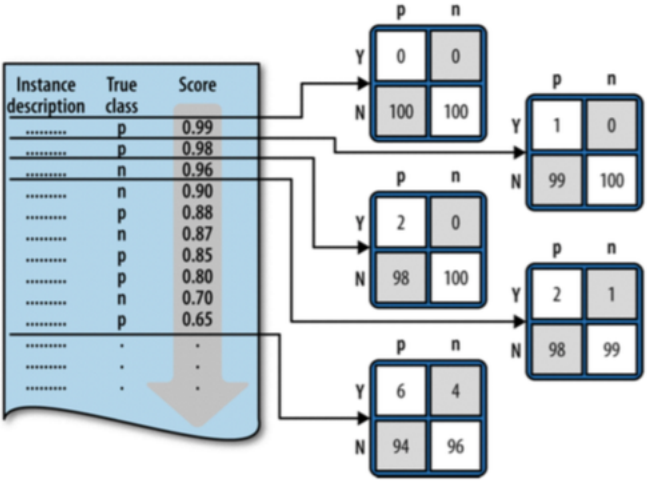
For a given model f, each threshold value T gives a point on the Profit Curve

Model score is the threshold probability classifying $+$ vs $-$

1. Allow T to be the maximum score
2. $TP = 0, FP = 0$
3. Calculate $E[Profit]$
4. For each observation, i:
   - If $\hat{\pi}_i > T \longrightarrow$ increment TP
   - Else $\longrightarrow$ increment FP
5. Add point (% Test Instances predicted Positive, $E[Profit]$) to the Profit Graph

Increment T from max-score to min-score, repeating steps 1-5

Building the Profit Curve



Example: Profit Curves for Multiple Classifiers