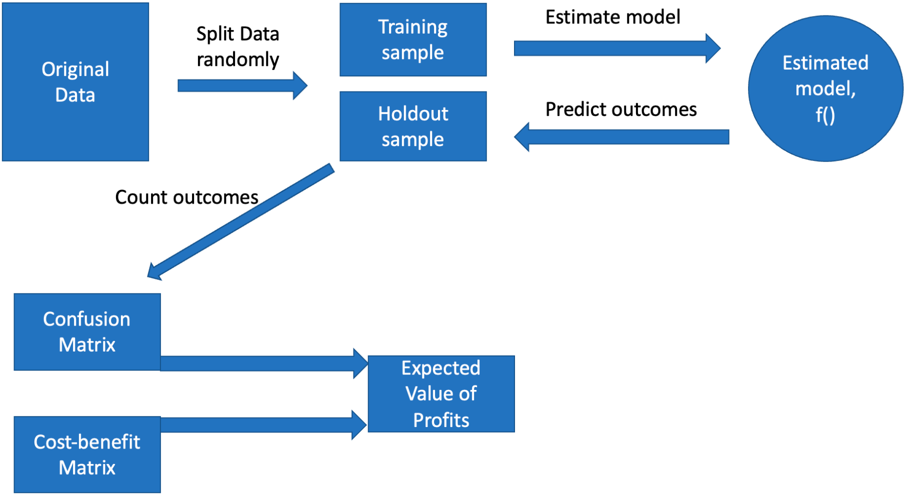
AI in Marketing: Prediction Validation

Main Steps to Evaluate Potential Profits from Prediction



Holdout Samples

How to measure the quality of predictions

- Randomly split your data set in two (50:50 is common, but other portions are used as well)
- Estimate your model on the **training sample**
  - Other names: "estimation," "learning" or "calibration" sample
- Test the quality of your predictions on the **holdout sample**
  - Other names: "validation," "leave out" or "test" sample
- This measures the quality of your predictions on *data your model has not seen before.*
- This mimics how your model will be used: *on outcomes you cannot see.*
- This tests how well your prediction model *generalizes* to new data from the same source as the training data.
- Using the same data to train and validate your model is "cheating"

Estimate model

- Estimate the model on the "estimation sample"
- Various models can be used at this step.
- The estimated has specific numbers for each parameter in the model (e.g., estimated coefficients)

For example, a logit model:

$$\text{Prob[outcome]} = \frac{\exp\left(\beta_0 + \beta_1 X_1 + \beta_2 X_2\right)}{1 + \exp\left(\beta_0 + \beta_1 X_1 + \beta_2 X_2\right)}$$

When you estimate the model you get specific values for each β. We refer to the estimated values as $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$.

Estimate model

- Estimate the model on the "estimation sample"
- Various models can be used at this step.
- The estimated has specific numbers for each parameter in the model (e.g., estimated coefficients)

For example, a logit model:

$$\text{Prob}[\text{outcome}] = \frac{\exp\left(\beta_0 + \beta_1 X_1 + \beta_2 X_2\right)}{1 + \exp\left(\beta_0 + \beta_1 X_1 + \beta_2 X_2\right)}$$

When you estimate the model you get specific values for each β. We refer to the estimated values as $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$.

Make Predictions using Model

- The estimated model allows you to make predictions.
- Predictions should be made on the **holdout sample**
  - See how well the model generalizes to new data sets
- For a logit model the output will be probabilities, $p \in [0, 1]$
- To make the predictions outcomes, $o \in \{0, 1\}$, we need to pick a threshold, s. Assign
  - Prediction $= 0$ if $p < s$
  - Prediction $= 1$ if $p > s$
- The threshold choice is something we can experiment with to improve expected profits.
  - More on this later.

Make Predictions using Model

- The estimated model allows you to make predictions.
- Predictions should be made on the **holdout sample**
  - See how well the model generalizes to new data sets
- For a logit model the output will be probabilities, $p \in [0, 1]$
- To make the predictions outcomes, $o \in \{0, 1\}$, we need to pick a threshold, s. Assign
  - Prediction $= 0$ if $p < s$
  - Prediction $= 1$ if $p > s$
- The threshold choice is something we can experiment with to improve expected profits.
  - More on this later.

Compute the Confusion Matrix

We now have data in the holdout sample that look like this:

Predicted Outcome	Actual Outcome
1	0
1	1
0	0
0	1
0	0

To get the confusion matrix we just have to count the total number of each event.

	Predicted 0	Predicted 1
Actual 0	N_1	N_3
Actual 1	N_2	N_4

Calculate Expected Value

Confusion Matrix Probabilities

	Predicted 0	Predicted 1
Actual 0	$p(o_1)$	$p(o_3)$
Actual 1	$p(o_2)$	$p(o_4)$

Cost-Benefit Matrix

	Predicted 0	Predicted 1
Actual 0	$v(o_1)$	$v(o_3)$
Actual 1	$v(o_2)$	$v(o_4)$

Expected Value (EV) of the Model's Predictions

$$EV = p(o_1) \times v(o_1) + p(o_2) \times v(o_2) + p(o_3) \times v(o_3) + p(o_4) \times v(o_4)$$

How to Pick a Threshold

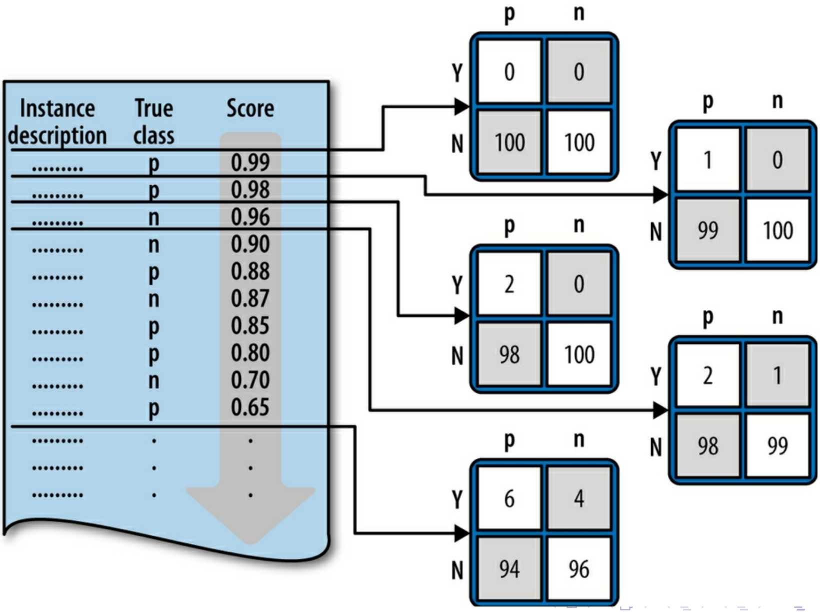
- Recall that the estimated model predicts probabilities for each outcome:
$$\text{Prob[outcome]} = \frac{\exp\left(\beta_0 + \beta_1 X_1 + \beta_2 X_2\right)}{1 + \exp\left(\beta_0 + \beta_1 X_1 + \beta_2 X_2\right)}$$
- These predictions are also known as the "**score**."
- The **scores** allow you to rank which consumers are most attractive
  - Spend your marketing effort on the highest **score** customers first.

How to Pick a Threshold

- Order your data by score, from highest to lowest.
- Each observation in the data has a different confusion matrix associated with it.
  - Calculate the number of true positives and false positives assuming each score level is the threshold, s.
- Each confusion matrix has a different expected profit associated with it.
- You can pick the score threshold that gives the highest expected profits.

How to Pick a Threshold

Data Science for Business, Figure 8-1

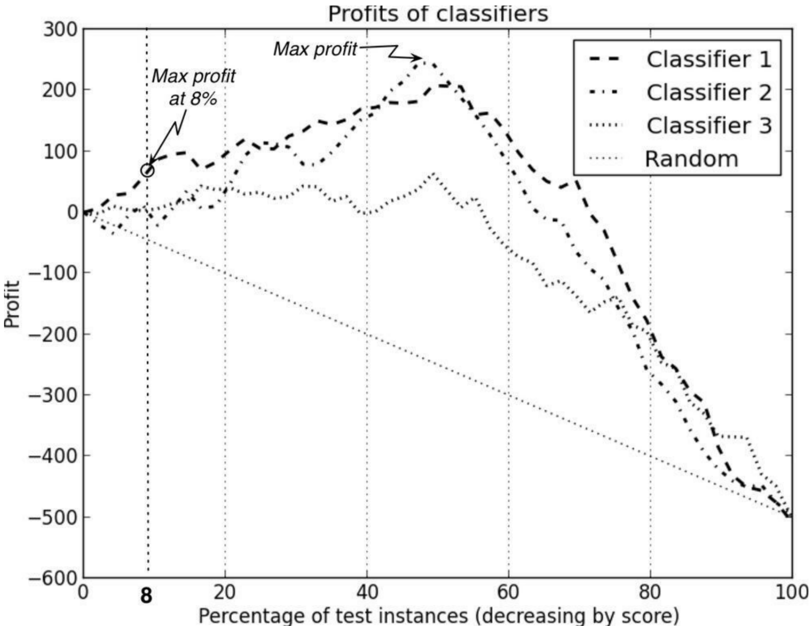


How to Pick a Threshold

- You could plot the expected profits associated with each threshold level.
- A very similar, but more helpful plot is the expected profits associated with targeting some percentage of the population.
  - This allows easier comparison across models
- This plot is called the **Profit Curve**

Profit Curve

Data Science for Business, Fig 8-2



Profit Curve

- Allows comparison of models when there is a limited advertising budget.
- The example illustrated on the previous graph assumes there is only enough money to advertise to 8% of your customers
- In this case, Classifier 1 is the best choice.
  - See "Max profit at 8%" arrow.
- However, with unlimited budget, Classifier 2 is the best choice.
  - See "Max Profit" arrow.

Profit Curve: Technical Note

Receiver Operating Characterstic (ROC) curve

- The Profit Curve is closely related to Receiver Operating Characteristic curve (ROC).
- Both are measure of the quality of binary predictions.
- The Profit Curve has direct implications for business decicions, ROC doesn't.
- ROC does not require a cost-benefit matrix and is often used by statisticians.
- ROC is often summarized with a single number called "Area Under the Curve" (AUC)
    - Alternative measure of prediction quality, different from expected profits.

Ways to Improve Expected Value of Profits

- Experiment with the probability threshold, s.
  - You can try out many different values to see which gives the highest EV.
- Experiment with different models.
  - No machine learning model is best in all situations. Try several and see what works best for your data.
- Experiment with different tuning parameters.
  - More on this later in the course.
  - Tuning affects some algorithms more than others.
- Get more data.
  - Additional observations typically improve predictions. Typically you need many times more observations to see a meaningful improvement.
  - Additional variables that are predictive of outcomes can improve predictions and therefore EV.

Polling questions

Access the poll here:

www.pollEv.com/mthomas

- Log in with your NUS ID
- This will allow me to give credit for your participation.

Practice Question

Which of the following is *most immediately* used to produce the confusion matrix?

- Original data
- Training sample
- Ans: Holdout sample
- Profit curves.
- Model parameters.

Practice Question

The difference between the training and the holdout sample is:

- The highest-quality observations are included in the training data.
- Ans: There is no statistical difference between them, in expectation.
- The training data include the score values.
- The holdout data are used to estimate model parameters.
- The training data report the expected profits for each observation.

Practice Question

Which of the following is true of a model's "score" values

- They typically take values 0 or 1.
- Counting their values allows us to produce the confusion matrix.
- They report the estimated parameters of the model
- They are most useful when included in the training data
- Ans: They help rank customers' attractiveness.

Practice Question

The Profit Curve reports

- Profits expected for each threshold level.
- Ans: Profits expected by targeting some percentage of the population.
- Profits expected for each score level.
- Profits expected from the best model only.
- Profits from people in the training data.

Practice Question

Why does the profit curve eventually go down (as you move left to right)?

- Ans: The quality of customers to target eventually declines.
- The quality of the prediction quality eventually declines.
- The marketing cost keeps increasing.
- The precision of our estimates eventually declines.
- The out-of-sample prediction error increases.