Data Science for Business
Lecture #7
Evaluating Profits for Freemium

Prof. Alan L. Montgomery

Carnegie Mellon University, Tepper School of Business

email: alanmontgomery@cmu.edu

Lecture Outline

How do we know if our model is a good one?

Deciding upon the best model using decision making

Computing the profitability of an free promotional offer for freemium

Decision Tree Analysis

Which decision tree is best?

How to compare models?

What is the appropriate sample to use for comparing our models?

Training sample

- Use for letting the algorithm determine parameter values

Validation sample

- Use for determining the “best” model

Prediction sample

- Use to determine accuracy of the “best” model

Model Comparison across Trees

	Simple tree	Complex tree
Accuracy in Validation	91.9%	92.6%
Precision in Validation	31.8%	35.9%
Lift in top decile in validation	2.23	3.09
AUC	0.697	0.697

Very complex (cp=.0001) Training	Very complex (cp=.0001) Validation
93.0%	88.0%
48.3%	19.8%

Which model should we use?

Are we overfitting?

Revisit Freemium Exercise

What is our payoff matrix for Freemium?

		Prediction	
		Do not subscribe	Subscribe
Action	Do not subscribe		
	Subscribe		

Discussion Exercise

Evaluate several different cut-offs for the Freemium Logistic Regression. What happens if we predict “adoption” if probability is…

- Greater than 30%
- Greater than 50%
- Greater than 90%

Which is the right cut-off?

Evaluating the Tree Model

```
> # predict probability (for validation sample)
> padopter = predict(ctree,newdata=rfreemium[validsample,crvarlist],type='vector')
> cadopter = (padopter>.25)+0    # notice that we use a cutoff of 25% because it is harder to predict adopters
> trueadopter = freemium$adopter[validsample]
> (results = xtabs(~cadopter+trueadopter) )  # confusion matrix (columns have truth, rows have predictions)
        trueadopter
cadopter     0     1
       0 29438  1876
       1   500   280
> (accuracy = (results[1,1]+results[2,2])/sum(results) )  # how many correct guesses along the diagonal
[1] 0.9259675
> (truepos = results[2,2]/(results[1,2]+results[2,2]))  # how many correct "adopter" guesses
[1] 0.1298701
> (precision = results[2,2]/(results[2,1]+results[2,2])) # proportion of correct positive guesses
[1] 0.3589744
> (trueneg = results[1,1]/(results[2,1]+results[1,1]))  # how many correct "non-adopter" guesses
[1] 0.9832988
> # compute the predictions for the 10% of most likely adopterers (for validation sample)
> topadopter = as.vector(padopter>=as.numeric(quantile(padopter,probs=.9)))
> ( baseconv=sum(trueadopter==1)/length(trueadopter) )  # what proportion would we have expected purely due to chance
[1] 0.06717767
> ( actconv=sum(trueadopter[topadopter])/sum(topadopter))  # what proportion did we actually predict
[1] 0.2076367
> ( lift=actconv/baseconv )  # what is the ratio of how many we got to what we expected
[1] 3.09086
```

Evaluating the Logistic Regression Model

```
> # predict probability (for validation sample)
> padopter = predict(fwd,newdata=rfreemium[validsample,crvarlist],type='response')
> cadopter = (padopter>.25)+0    # notice that we use a cutoff of 25% because it is harder to predict adopters
> trueadopter = freemium$adopter[validsample]
> (results = xtabs(~cadopter+trueadopter) )  # confusion matrix (columns have truth, rows have predictions)
        trueadopter
cadopter     0     1
       0 29576  1958
       1   362   198
> (accuracy = (results[1,1]+results[2,2])/sum(results) )  # how many correct guesses along the diagonal
[1] 0.9277123
> (truepos = results[2,2]/(results[1,2]+results[2,2]))  # how many correct "adopter" guesses
[1] 0.09183673
> (precision = results[2,2]/(results[2,1]+results[2,2])) # proportion of correct positive guesses
[1] 0.3535714
> (trueneg = results[1,1]/(results[2,1]+results[1,1]))  # how many correct "non-adopter" guesses
[1] 0.9879083
> # compute the predictions for the 10% of most likely adopterers (for validation sample)
> topadopter = as.vector(padopter>=as.numeric(quantile(padopter,probs=.9)))
> ( baseconv=sum(trueadopter==1)/length(trueadopter) )  # what proportion would we have expected purely due to chance
[1] 0.06717767
> ( actconv=sum(trueadopter[topadopter])/sum(topadopter))  # what proportion did we actually predict
[1] 0.2214953
> ( lift=actconv/baseconv )  # what is the ratio of how many we got to what we expected
[1] 3.297157
> # predict probability (for prediction sample)
> padopter = predict(fwd,newdata=rfreemium[predsample,crvarlist],type='response')
> cadopter = as.vector((padopter>.25)+0)  # classify the predictions as adopters or not
> trueadopter = freemium$adopter[predsample]
> (results = xtabs(~cadopter+trueadopter))  # confusion matrix
        trueadopter
cadopter    0    1
       0 9799  659
       1  123   71
> (accuracy = (results[1,1]+results[2,2])/sum(results) )  # how many correct guesses along the diagonal
[1] 0.9265866
> (truepos = results[2,2]/(results[1,2]+results[2,2]))  # how many correct "adopter" guesses
[1] 0.09726027
> (precision = results[2,2]/(results[2,1]+results[2,2])) # proportion of correct positive guesses
[1] 0.3659794
> (trueneg = results[1,1]/(results[2,1]+results[1,1]))  # how many correct "non-adopter" guesses
[1] 0.9876033
```

Model Accuracy (Validation Sample) using different cut-offs

Cut-off	# Pos	Accuracy	TruePos	Precision	TrueNeg
0.05	16240	53.5%	80.7%	10.7%	51.6%
0.10	3514	87.1%	35.2%	21.6%	90.8%
0.25	560	92.8%	9.2%	35.4%	98.8%
0.50	154	93.1%	2.6%	35.7%	99.7%
0.90	28	93.2%	3.2%	25.0%	99.9%

Which cut-off is best?

What is our payoff matrix?
If we do nothing

		Prediction	
		Subscribe	Do Not Subscribe
Action	Subscribe		
	Do Not Subscribe		

What is our payoff matrix?
If we do nothing

		Prediction	
		Subscribe	Do Not Subscribe
Action	Subscribe	Fees= $3x12=$36	Fees= $3x12=$36
	Do Not Subscribe	Adv =$0.125x12 =$1.50	Adv =$0.125x12 =$1.50

Who to make offers to?

Payoff from Offer

		Prediction	
		Subscribe	Do Not Subscribe
Action	Subscribe		
	Do Not Subscribe		

Payoff from No Offer

		Prediction	
		Subscribe	Do Not Subscribe
Action	Subscribe	Fees= $3x12=$36	Fees= $3x12=$36
	Do Not Subscribe	Adv =$0.125x12 =$1.50	Adv =$0.125x12 =$1.50

Assume no dropout (e.g., will stay subscriber the entire 12 months) or new users (e.g., add a friend)

Who to make offers to?

Payoff from Offer

		Prediction: Subscribe	Prediction: Do Not Subscribe
Action	Subscribe	Fees-Opportunity =\$3x12-\$3x3 =\$27	Fees= \$3x12=\$36
Action	Do Not Subscribe	Adv-Opportunity =\$0.125x12-\$0.125x3 =\$0.125x9 =\$1.125	Adv =\$0.125x12 =\$1.50

Payoff from No Offer

		Prediction: Subscribe	Prediction: Do Not Subscribe
Action	Subscribe	Fees= \$3x12=\$36	Fees= \$3x12=\$36
Action	Do Not Subscribe	Adv =\$0.125x12 =\$1.50	Adv =\$0.125x12 =\$1.50

Assume no dropout (e.g., will stay subscriber the entire 12 months) or new users (e.g., add a friend)

What is the expected payoff?

Payoff from Offer

$$
\begin{aligned}
&\mathrm{E}[\mathit{Profit} \mid \mathit{Offer}] = \\
&\ \$27 \cdot \Pr(\mathit{Adopt} \wedge \mathit{Offer}) \\
&+\$36 \cdot \Pr(\mathit{Adopt} \wedge \mathit{NoOffer}) \\
&+\$1.25 \cdot \Pr(\mathit{NotAdopt} \wedge \mathit{Offer}) \\
&+\$1.50 \cdot \Pr(\mathit{NotAdopt} \wedge \mathit{NoOffer}) \\
&= \$27 \cdot p' \cdot I(p' \geq \tau) \\
&+\$36 \cdot p' \cdot I(p' < \tau) \\
&+\$1.25 \cdot (1-p') \cdot I(p' \geq \tau) \\
&+\$1.50 \cdot (1-p') \cdot I(p' < \tau) \\
&= (\$25.75 p' + \$1.25) \cdot I(p' \geq \tau) \\
&+(\$34.50 p' + \$1.50) \cdot I(p' < \tau)
\end{aligned}
$$

Payoff from No Offer

$$
\begin{aligned}
&\mathrm{E}[\mathit{Profit} \mid \mathit{NoOffer}] = \\
&\ \$36 \cdot \Pr(\mathit{Adopt} \wedge \mathit{Predict}) \\
&+\$36 \cdot \Pr(\mathit{Adopt} \wedge \mathit{NoPredict}) \\
&+\$1.50 \cdot \Pr(\mathit{NotAdopt} \wedge \mathit{Predict}) \\
&+\$1.50 \cdot \Pr(\mathit{NotAdopt} \wedge \mathit{NoPredict}) \\
&= \$36 \cdot p \cdot I(p \geq \tau) \\
&+\$36 \cdot p \cdot I(p < \tau) \\
&+\$1.50 \cdot (1-p) \cdot I(p \geq \tau) \\
&+\$1.50 \cdot (1-p) \cdot I(p < \tau) \\
&= \$36 \cdot p + \$1.50 \cdot (1-p)
\end{aligned}
$$

where $p' = \Pr(\mathit{Adopt} \mid \mathit{Offer})$, $p = \Pr(\mathit{Adopt} \mid \mathit{NoOffer})$,

$\tau = \mathit{Cutoff}$ to classify, $I(p \geq \tau) = 1$ if $p \geq \tau$ and $=0$ if $p < \tau$

The profitability of making offers

The incremental profitability of making an offer versus making no offers:

$$
\begin{aligned}
&\mathrm{E}[\mathit{Incremental\ Profit\ from\ Offer}] \\
&= \mathrm{E}[\mathit{Profit} \mid \mathit{Offer}] - \mathrm{E}[\mathit{Profit} \mid \mathit{NoOffer}] \\
&= \left\{(\$25.75p' + \$1.25)\cdot I(p' \geq \tau) + (\$34.50p' + \$1.50)\cdot I(p' < \tau)\right\} - \left\{\$36\cdot p + \$1.50\cdot(1-p)\right\} \\
&= \left\{(\$27\cdot p' + \$1.25\cdot(1-p') + (\$8.75p' + \$0.25)\cdot I(p' < \tau)\right\} - \left\{\$36\cdot p + \$1.50\cdot(1-p)\right\} \\
&= \underbrace{\left\{(\$27\cdot p' + \$1.25\cdot(1-p')\right\}}_{\text{Base profits with offers}} - \underbrace{\left\{\$36\cdot p + \$1.50\cdot(1-p)\right\}}_{\text{Base profits without offers}} + \underbrace{\left(\$9\cdot p' + \$0.25(1-p')\right)\cdot I(p' < \tau)}_{\text{Increment from promoting to those that would not have adopted}}
\end{aligned}
$$

Notice that if we make offers to everyone then incremental profits is 0. The increased profitability is determined by our targeting of *convincible* non-adopters and the increase in probability that comes from our offers.

How does an offer effect the probability of adopting premium?

We can use either a logistic regression or a decision tree to make predictions about whether a customer will adopt:

$$\Pr(Adopt \mid Characteristics) = p$$

Unfortunately, these predictions do not take into account the actions that we will take. What we want is the conditional probability of churn given all the characteristics, but also what is the probability of adopt given a new offer (free three month subscription):

$$\Pr(Adopt \mid Characteristics, Offer) = p'$$

We are using a Data approach but we have never seen the target!

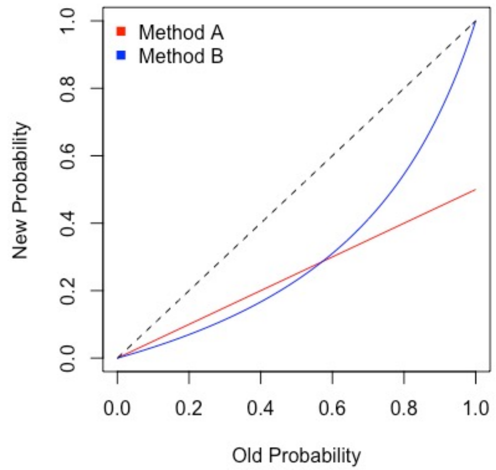
Adjusting Adoption Probability given Promotional Offer: Two methods for adjusting churn probability for Offers

Method A. Assume probability is multiplied by factor.

$$p' = \alpha \cdot p \quad \Rightarrow \quad \frac{p'}{1-p'} = \frac{\alpha p}{1-\alpha p} = \left(\alpha \frac{1-p}{1-\alpha p}\right)\left(\frac{p}{1-p}\right)$$

Method B. Assume that the odds ratio is multiplied by factor.

$$\frac{p'}{1-p'} = \beta \cdot \frac{p}{1-p} \quad \Rightarrow \quad p' = \frac{\beta}{1-p+p\beta}p$$



Comments

- Both methods are correct
- Work with predictions from Tree or LR
- The challenge: factor is unknown

Adjusting Adoption Probability given Promotional Offer: *Understanding Method B as new variable in logistic regression*

You have estimated the following logistic regression model:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

However, we need a model that shows how the probability will change if we make an offer so we can add an offer term to model:

$$\ln\left(\frac{p'}{1-p'}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \gamma \cdot Offer$$

The "new" model is just the same as the first model with a new offer effect:

$$\frac{p'}{1-p'} = \frac{p}{1-p} \cdot \beta \quad \Rightarrow \quad \ln\left(\frac{p'}{1-p'}\right) = \ln\left(\frac{p}{1-p}\right) + \gamma \cdot Offer$$

- Notice we can add "offer term" to the model, but we cannot estimate it since we do not have any data on this variable.

What is the effect of a promotion?
Using Method B

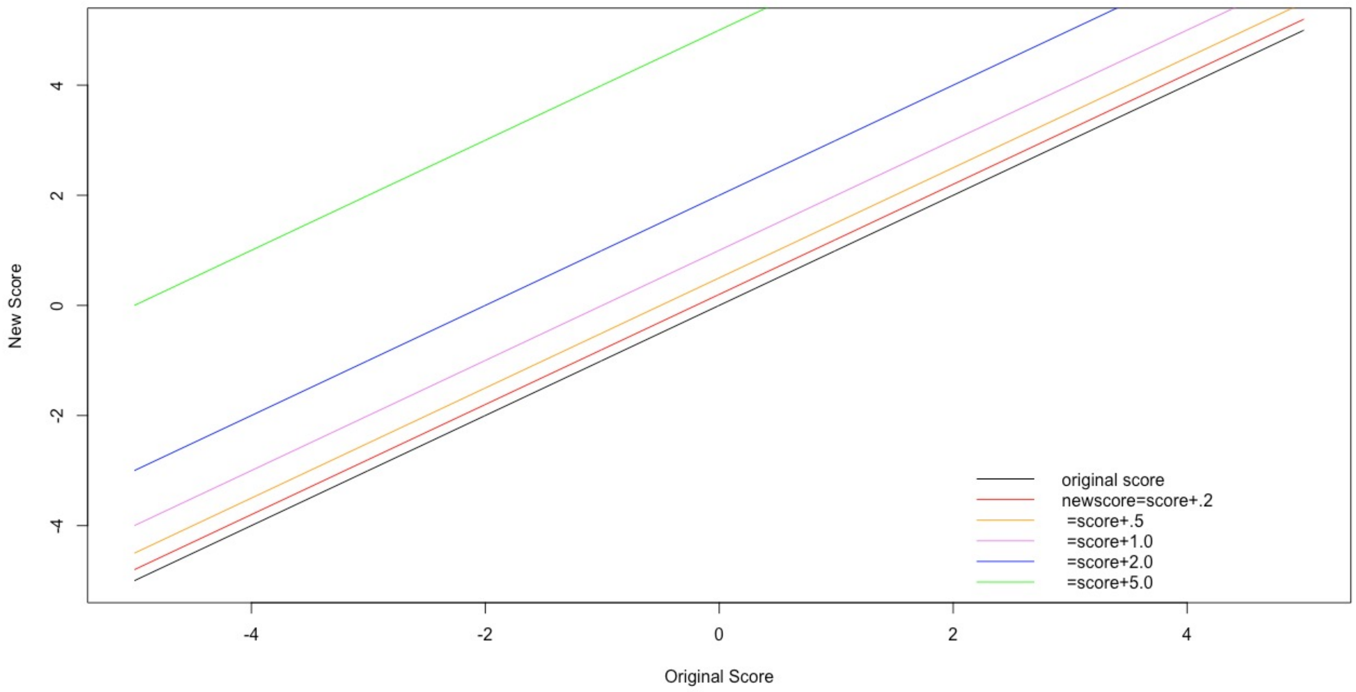
What is the correct values of $\gamma \cdot Offer$:

$$\frac{p'}{1-p'} = \frac{p}{1-p} \cdot \beta \quad \Rightarrow \quad \ln\left(\frac{p'}{1-p'}\right) = \ln\left(\frac{p}{1-p}\right) + \gamma \cdot Offer$$

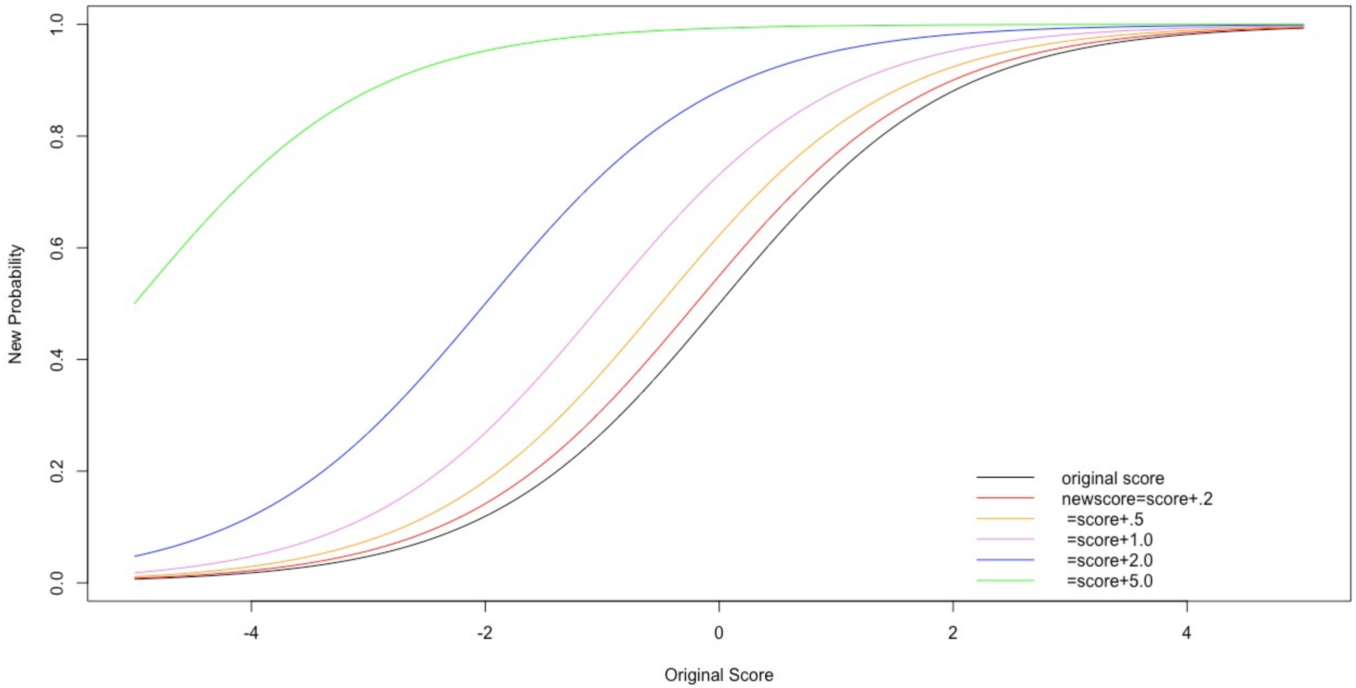
How could we figure this out?

- Experience
  - Perhaps with previous customers, other companies, other regions
- Experimentation
  - But this takes time and can be expensive
- Guess
  - What do you think a reasonable guess of giving a 3 month promotional offer?
  - Will it turn a customer with conversion probability at 1% into 1.1%? 2%? 10%?

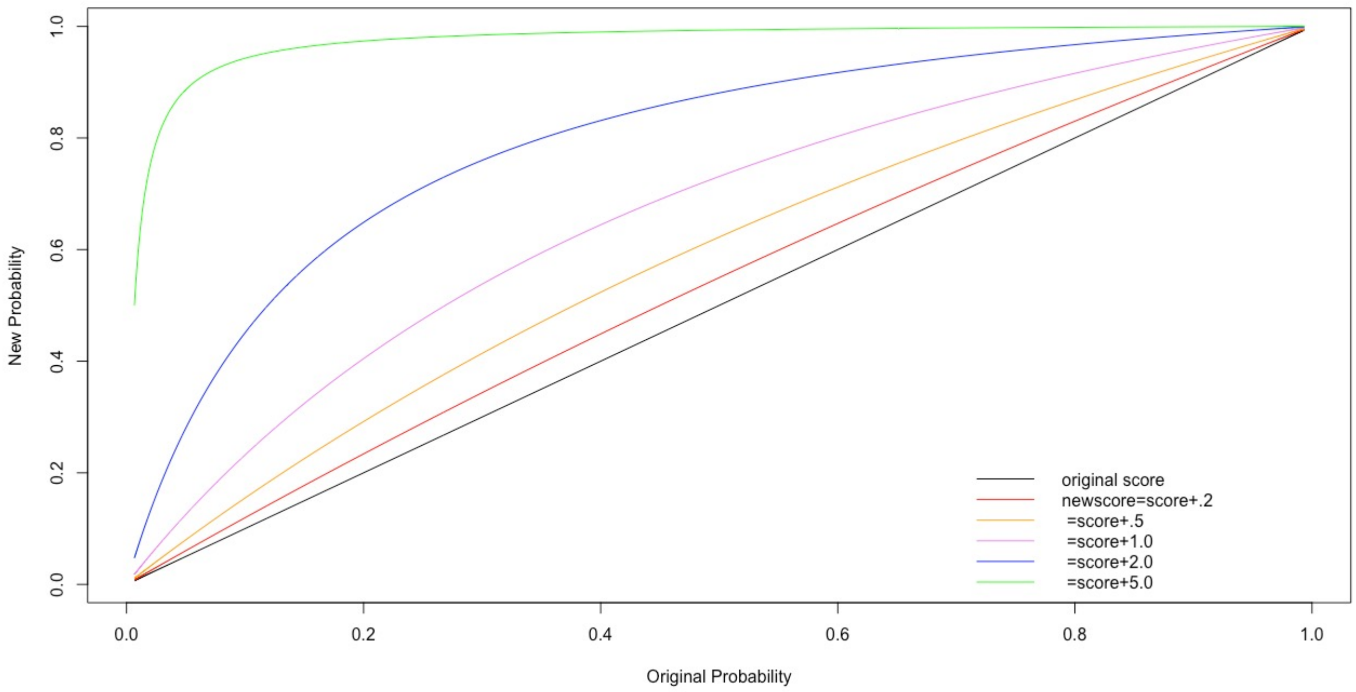
Modify the score for different $\gamma \cdot Offer$



Modify the score for different $\gamma \cdot Offer$



Modify the score for different $\gamma \cdot Offer$



What is payoff from no offer?

Payoff from No Offer

		Prediction	
		Subscribe	Do Not Subscribe
Action	Subscribe	Fees= \$3x12=\$36	Fees= \$3x12=\$36
	Do Not Subscribe	Adv =\$0.125x12 =\$1.50	Adv =\$0.125x12 =\$1.50

E[Profit|No Offer]

=

Prob(Subscribe|No Offer) x (Profits of Subscribe without Offer)

+ Prob(Do Not Subscribe|No Offer) x (Profits of Do Not Subscribe without Offer)

= p x \$36 + (1-p) x \$1.50

Assume no dropout (e.g., will stay subscriber the entire 12 months) or new users (e.g., add a friend)

What is payoff from offer

Payoff from Offer

		Prediction: Subscribe	Prediction: Do Not Subscribe
Action	Subscribe	Fees-Opportunity =\$3x12-\$3x3 =\$27	Fees= \$3x12=\$36
Action	Do Not Subscribe	Adv-Opportunity =\$0.125x12-\$0.125x3 =\$0.125x9 =\$1.125	Adv =\$0.125x12 =\$1.50

Suppose we give offer to everyone:

E[Profit|Offer]

=

Prob(Subscribe|Offer) x (Profits of Subscribe with Offer)

+ Prob(Do Not Subscribe|Offer) x (Profits of Subscribe with Offer)

= p’ x \$27 + (1-p’) x \$1.125

Assume no dropout (e.g., will stay subscriber the entire 12 months) or new users (e.g., add a friend)

What is the payoff from an offer?

Gain = E[Profit|Offer] - E[Profit|No Offer]

= { Prob(Subscribe|Offer) x (Profits of Subscribe with Offer)

+ Prob(Do Not Subscribe|Offer) x (Profits of Subscribe with Offer) }

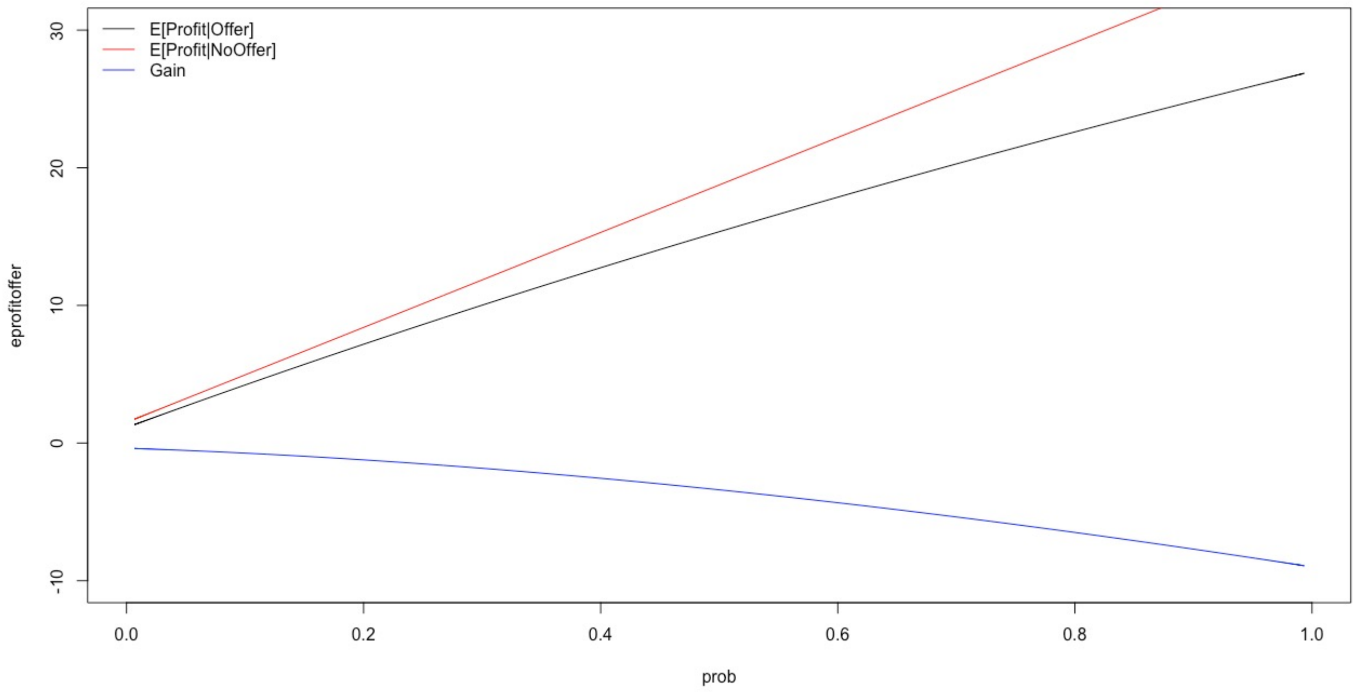
- { Prob(Subscribe|No Offer) x (Profits of Subscribe without Offer)

+ Prob(Do Not Subscribe|No Offer) x (Profits of Do Not Subscribe without Offer) }

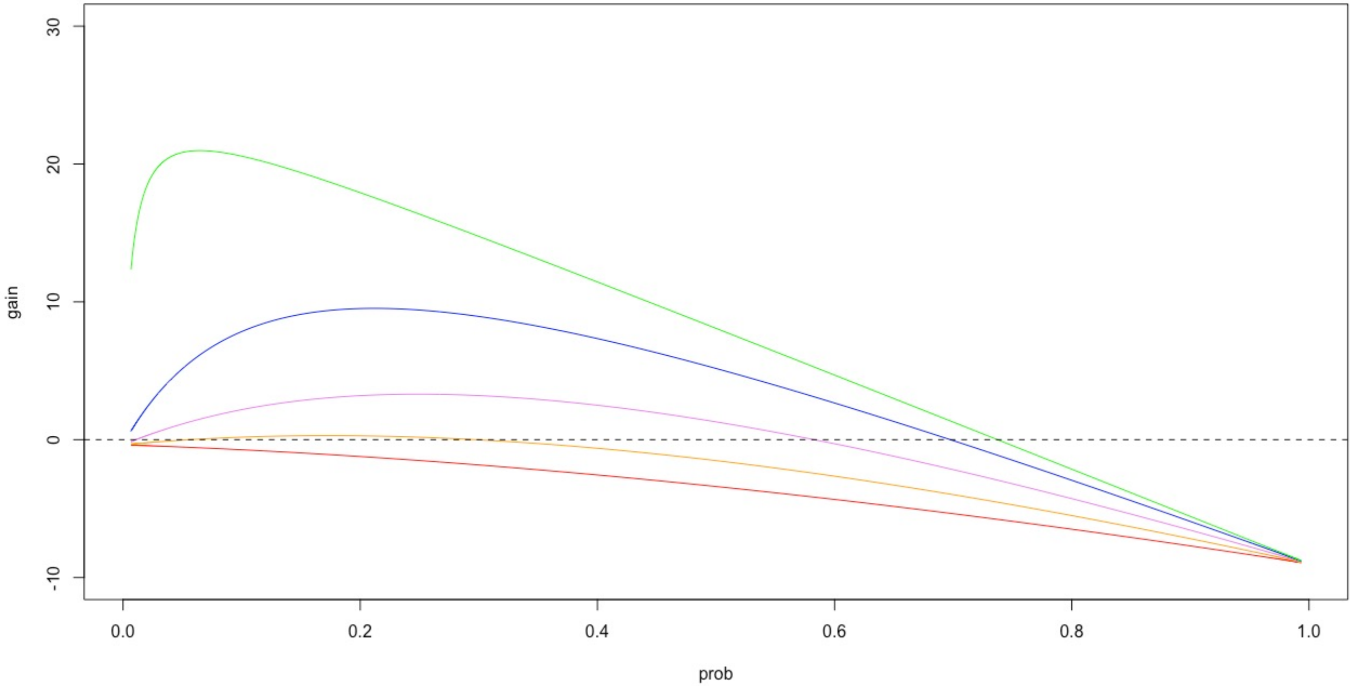
= { p’ x $27 + (1-p’) x $1.125 } – { p x $36 + (1-p) x $1.50 }

When is this positive?

What happens if $\gamma \cdot Offer = 0.2$?



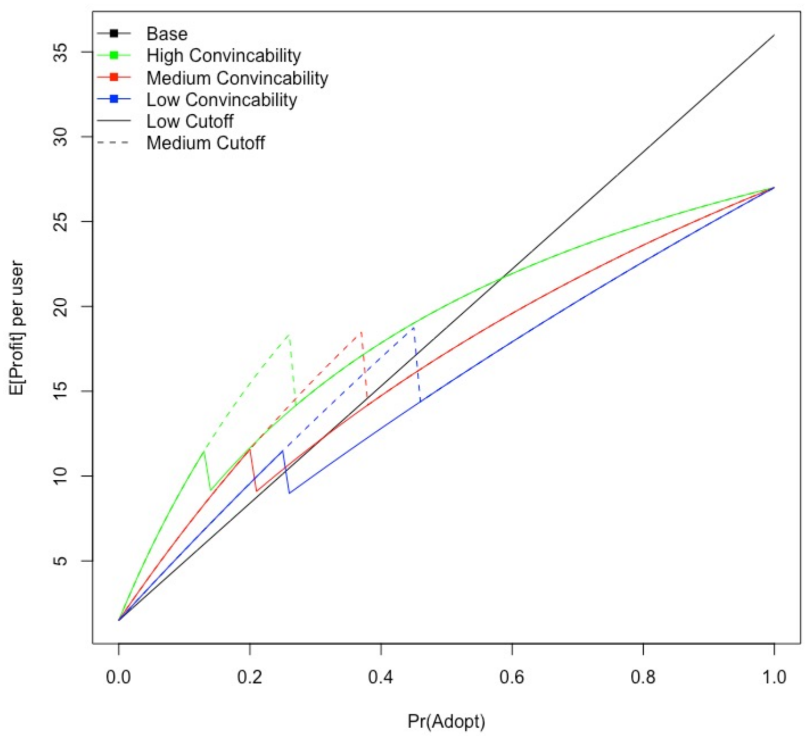
What about other values?



What is the meaning of the dashed line?

Who should we give offers to?

Gain in E[Profits] as a function of Convincability and Cut-off Values



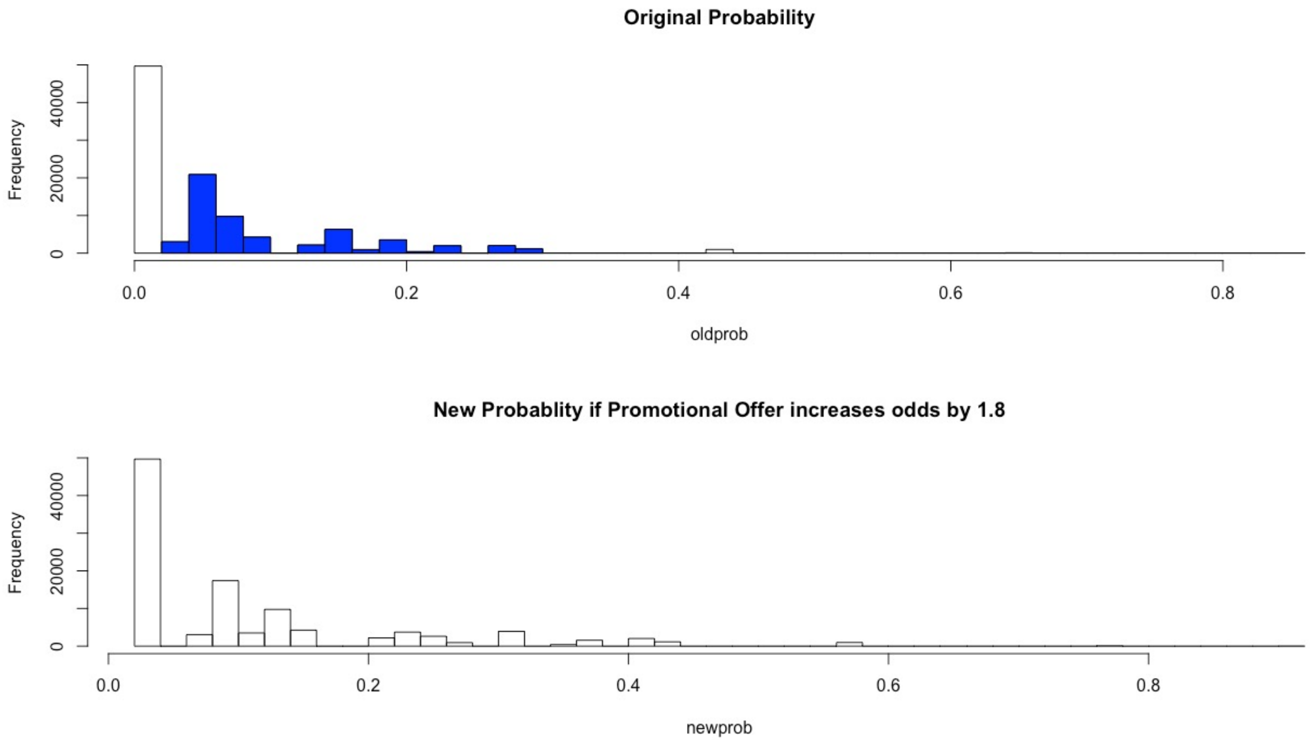
Base Profits = 36 x p + 1.5 * (1-p)

New Profits = 27 x pprime + 1.25 x (1-pprime)
+ [(9 x pprime) + 0.25 x (1-pprime)]
x I(pprime<cutoff)

Notice that our profits are higher when customers are easier to convince (e.g., offers work well) and when it is easier to classify (e.g., identify who to give offer to)

Profits decrease from promotional offers if the probability of customers adopter is already hig

Suppose Promotional Offer increases odds by 1.8
Using our complex decision tree



We would have expected to make $408,768 over the next year from our 100k users.

We should make offers to 56,454 consumers (anyone who's probability is more than 1.482%)

We increase our profits by 5.4% or an additional $22,165.

Conclusions

Our predictive model helps us understand who to give offers to

- However, the critical piece of information that we want to know is what is the impact of a free promotional offer on subscription
- But we don't know this from the model or the data. Why?

Using our knowledge about expected profits we realize:

- Give offers to users with relatively low probabilities of adopting
  - We only lose $.375 in advertising from the promotional period, but gain $27 in premiums
  - So there is a potential for big gains
- Don't give offers to users with high probabilities of adopting
  - We lose $9 when we give promotional offers to someone that subscribes without the promotional
  - So there is not a lot of upside to giving promotional offers to likely subscribers

Key Takeaways

Evaluate models based upon their impact on the outcome

- Usually more accurate models lead to better outcomes, but we care a lot about the costs of mistakes that we make

Choose the model that gives us the best result not the most accurate prediction

Often our data gives us the result if we do nothing, but we are interested in assessing what happens if we do something new

- "Something new" is not in our data. You may have to adjust your probabilities.
- Ideally run an experiment to learn the adjustment, but initially you need to make a guess. Evaluate the sensitivity of your recommendation to this guess.