Q -quantity sold,

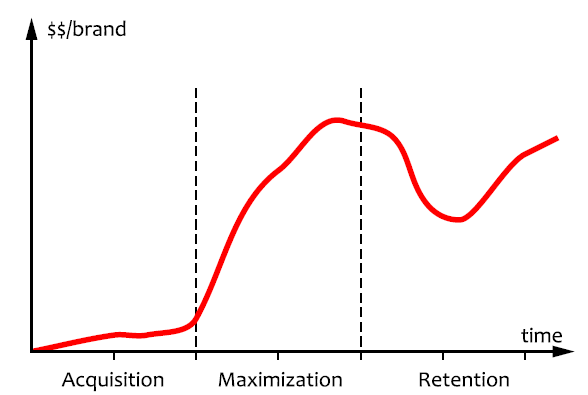
P - baseline unit price

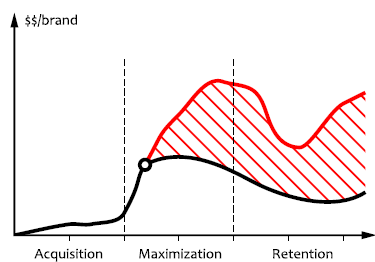
V – stands for variable campaign costs per unit   
C - fixed cost of a campaign

Qc - exceeds the sales volume without

the campaign   
- Q0 to an extent that is sufficient to cover the campaign

costs





.

P is the entire population of consumers,

U is the subset of consumers reached in the scope of the campaign,

GpUq is the expected profit of the campaign, which is a function of the targeting

strategy that selects U from P.

Recall that the basic marketing optimization problem is defined as

finding the strategy that maximizes the value function

This campaign is clearly inefficient or even

harmful. To understand this problem better, let us separately consider

the following four possible strategies:

1. Select a set of customers |U| according to equation 3.7 and send

promotions to everyone in this group

2. Select a set of customers |U| randomly and send promotions to

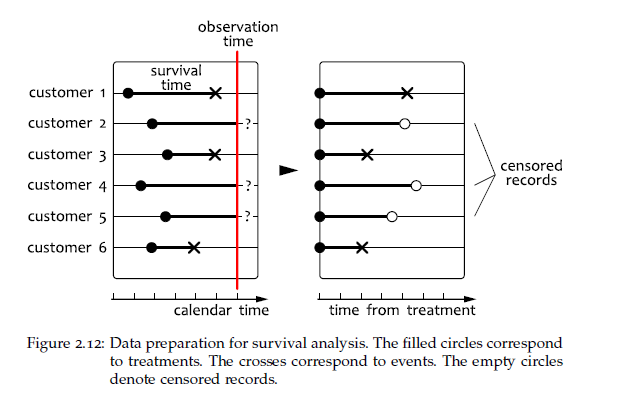
everyone in this group

3. Select a set of customers |U| according to equation 3.7 but do not

actually send the promotions

4. Select a set of customers |U| randomly but do not actually send

the promotions



Для оценки качества прогнозирования модели логистической регрессии в кредитном скоринге используют тест или статистику Колмогорова-Смирнова. В тесте КолмогороваСмирнова проверяется статистическая гипотеза, что две произвольные выборки принадлежат одной генеральной совокупности. Применительно к скорингу сравниваются два кумулятивных распределения скоринговых баллов «хороших» и «плохих» заемщиков. Статистика Колмогорова-Смирнова вычисляется как максимальная разница между кумулятивными функциями распределения «плохих» и «хороших» заемщиков

где и — эмпирические кумулятивные распределения скорингового балла для «плохих» и «хороших» заемщиков; n, m — количество «плохих» и «хороших» заемщиков.

Если расчетное значение статистики по формуле (3) меньше порогового по таблице или по формуле (4), то гипотезу о равенстве двух функций распределений отвергают. Диапазон изменения значений статистики Колмогорова-Смирнова может быть от 0 до 100. Чем выше значение статистики Колмогорова-Смирнова, тем лучше классифицирует модель. Крайние значения статистика Колмогорова-Смирнова, как правило, не принимает, и ее значения сосредоточены обычно от 20-25 до 75-80. Можно придерживаться следующей шкалы для оенки качества модели по статистике Колмогорова

После чего нужно найти максимальную разность между кумулятивным процентом «хороших» и «плохих» кредитов и вычислить по формуле (3) статистику КолмогороваСмирнова (см. рисунок 4). Полученное значение статистики необходимо сравнить с табличным значением по таблице распределения Колмогорова-Смирнова с выбранным уровнем значимости или при числе «плохих» и «хороших» заемщиков соответственно больше 80 можно взять приближенное пороговое значение, вычисляемое по формуле

Альтернативная мера оценки валидации модели является коэффициент дивергенции (от англ. Divergence). Коэффициент дивергенции представляет собой оценку разницы математических ожиданий распределений скоринговых баллов для «плохих» и «хороших» заемщиков, нормализованную дисперсиями этих распределений, и рассчитывается по формуле

Чем больше значение коэффициента дивергенции, тем лучше качество модели с точки зрения ее классификационной способности. В

|  |  |  |
| --- | --- | --- |
| **Распределение баллов** | **Статистика Колмогорова-Смирнова** | **Коэф.дивергенции** |
| Симметричны | Подходит | Подходит |
| Скошены внутрь | Подходит | Не походит (будет завышена) |
| Скошены наружу | Подходит | Не походит (будет завышена) |
| Вложены друг в груга | Не походит (будет завышена) | Подходит |

Значения Information Value (IV) для определения cutoff по отбору признаков:

* <0.02 Бесполезно для предсказания
* 0.02 – 0.1 Слабая
* 0.1 – 0.3 Средняя
* 0.3 – 0.5 Хорошая
* 0.5+ Слишком хорошо, что бы быть правдой

***Score = (β×WoE+ α/n)×Factor + Offset***

где:

β — коэффициент логистической регрессии признака

α — свободный член

WoE — Weight of Evidence признака

n — количество признаков, включенных в модель

Factor, Offset — параметры масштабирования. Множитель и смещение.

Множитель и смещение считаются так:

***Factor = pdo/Ln(2) Offset = B — (Factor × ln(Odds))***

где:

pdo — количество баллов, удваивающее шансы

B — значение на шкале баллов, в которой соотношение шансов составляет С:1