

Конспекты по ИПС

Минкин Даниэль

22 апреля 2025 г.

Содержание

1 Введение	1
2 Риск	2
3 Риск-менеджмент	2
3.1 Факторы, которые необходимо знать о риске	2
3.2 Риски и убытки	2
3.3 Что не входит в задачи риск-менеджмента?	3
3.4 Суть риск-менеджмента	3
4 Кредитный риск	3
4.1 Особенности кредитного риска	3
4.2 Компоненты кредитного риска	3
4.3 Эмпирические данные о распределении потерь портфеля	4
4.4 Мат. ожидание потерь	4
4.5 Right и Wrong way риски	4
5 ROC - кривая	5
5.1 TPR и FPR	5
5.2 Precision и Recall	6
5.3 Accuracy	6
6 ROC AUC	6
7 CAP	6

1 Введение

Мой ИПС посвящен построению скоринговой модели для предсказания потребительского дефолта. В рамках данного конспекта я собираю необходимую информацию из предоставленных куратором материалов

2 Риск

Риск — это влияние неопределенности на цели.

- Влияние - это отклонение от ожидаемого результата
- Цель - некий набор результатов (не обяз. с зад вероятностями)
- Неопределенность - это отсутствие информации о событии

Источник риска — сущность, которая самостоятельно или в комбинации с другими обладает возможностью вызывать повышение риска, i.e повышает вероятность того, что неопределенность повлияет на цель

Экспозиция — степень подверженности организации событию

Стейкхолдер (i.e причастная сторона) — сущность подвергающаяся воздействию риска i.e берущая на себя риск

3 Риск-менеджмент

3.1 Факторы, которые необходимо знать о риске

- Что является источником риска?
- Рисковое событие
- Какие последствия из тех, что произойдут при реализации риска нас интересуют?
- Экспозицию риска
- Степень возможности наступления события, выраженная как вероятность

3.2 Риски и убытки

Риск не связан с убытками, так как часть из них может быть достоверно определена и уже заложена при планировании

Убытки разделяются на два вида:

- **Ожидаемые** То, в чем мы уверены и заложили
- **Неожидаемые** Возможны, но не обязательны. Поэтому их необходимо оценить

3.3 Что не входит в задачи риск-менеджмента?

- Определение риска (это задачи аналитика или трейдера)
- Принятие решений

И.е риск менеджмент только просчитывает риски

3.4 Суть риск-менеджмента

- Риск не нужно минимизировать
- Риск-менеджмент определяет какие виды рисков нужно принимать, а какие нет
- Риск-менеджмент занимается неожиданными убытками
- Риск-менеджмент не предсказывает будущее, однако позволяет рассмотреть различные сценарии и дать им оценку
- Риск-менеджмент оценивает величину запасов

4 Кредитный риск

4.1 Особенности кредитного риска

- Асимметрия: у нас ограниченные доходы, но существенные убытки
- Распределение с тяжелыми хвостами для доходности
- Распределение доходностей не является нормальным
- Более длительный горизонт планирования
- Меньше наблюдаемых данных
- Дефолты редки, поэтому мало past data

4.2 Компоненты кредитного риска

- **EAD** (Exposure at Default) - экспозиция при дефолте
- **PD** (Probability of Default) - вероятность дефолта
- **LGD** (Loss Given Default) - сумма, которую мы получаем при дефолте (убытки или возможно прибыли). Определяется как относительная величина, а не абсолютная

Все компоненты определяются как случайные процессы, т.е эти параметры могут изменяться со временем

- EAD - сумма платежных обязательств, должника (за вычетом обеспечения) на момент дефолта
- Recovery rate - доля риска, которая может быть взыскана в ходе процедуры банкротства

$$LGD = (1 - recovRate) \quad (1)$$

где *recovRate* - recovery rate

Следовательно, в случае дефолта потери составят:

$$Loss_i = EAD * LGD \quad (2)$$

Общий loss скалывается из суммы loss - ов

Риск концентрации — ситуация когда индивидуальные потери внутри портфеля сильно коррелированы

4.3 Эмпирические данные о распределении потерь портфеля

- Асимметрия слева
- Тяжелые хвосты
- PD и EAD коррелированны
- LGD и EAD не коррелированны
- PD и LGD коррелированны
- Корреляции случайны

4.4 Мат. ожидание потерь

$$E[loss_i] = PD_i * EAD_i * LGD_i \quad (3)$$

При условии, что EAD_i и LGD_i независимы

Неожидаемый убыток — это разница между фактическим убытком и ожидаемым убытком

4.5 Right и Wrong way риски

Wrong-way риск — ситуация, при которой экспозиция контрагента положительно коррелирует с вероятностью дефолта этого контрагента

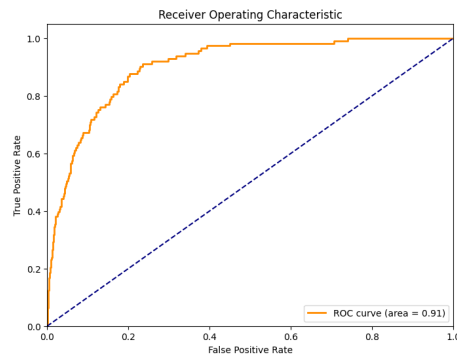
Если вероятность дефолта падает, но сумма, которую мы должны от него получить снижается

Right-way риск — ситуация, при которой EAD и PD коррелируют положительно

5 ROC - кривая

ROC (Receiver Operating Characteristic curve) — кривая операционной характеристики получателя, кривая на которой по оси Y расположен **true positive rate** или **чувствительность модели**, а по оси X **false positive rate** или **специфичность модели**

Предполагается, что у нас есть гиперпараметр, меняя который можно регулировать кол-во TP и FP. Данный параметр часто называется **точкой отсечения (cut-off value)** — некая величина p_{limit} такая, что если $p_{predicted} > p_{limit}$ то клиент считается дефолтным. Получается регулируя данный параметр мы можем менять оценку кол-ва дефолтных клиентов.



5.1 TPR и FPR

Таблица 1: Исходы предсказания

	Предсказано: Да	Предсказано: Нет
Фактически: Да	True Positive (TP)	False Negative (FN)
Фактически: Нет	False Positive (FP)	True Negative (TN)

$$TPR = \frac{TP}{TP + FN} \quad (4)$$

Intuition: Получается мы берем, все случаи когда произошел дефолт и смотрим, какую долю произошедших дефолтов мы предсказали

$$FPR = \frac{FP}{FP + TN} \quad (5)$$

Intuition: В этот раз мы смотрим на все случаи, когда дефолт не произошел и смотрим, долю ошибок нашей модели - когда мы сказали, что дефолт будет, но его не было

5.2 Precision и Recall

Precision и Recall - очень похожие метрики

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

В этот раз, мы рассматриваем все случаи, когда мы предсказали “Да” и оцениваем какая доля из этих предсказаний, оказалась правильной

$$Recall = TPR \quad (7)$$

5.3 Accuracy

Ассигасу используется для оценки общего качества модели, мы рассматриваем долю всех правильных ответов модели.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

6 ROC AUC

ROC AUC расшифровывается как ROC area under curve. Нам важна не самая кривая а ее форма. Поэтому для сравнительной оценки качества моделей, часто применяется сравнение площадей под ними. Легко заметить, что идеальная модель, будет иметь площадь под графиком равную единице, т.е кривая будет проходить по верхней границе $y = 1$ при любом x . (т.е мы будем предсказывать все случившиеся дефолты на 100% при этом во всех случаях, когда дефолта не было мы так и сказали). Также можно легко заметить, что если площадь под кривой менее 0.5, то наша модель хуже случайного классификатора, так как он представляет собой прямую $y = x$ (точка такой модели зависит от cut-off value)

Таким образом, с помощью AUC мы можем сравнивать качество моделей, однако данный показатель имеет мало смысла в абсолютном значении

На практике все определяется конкретными требованиями к модели, так модель с меньшим AUC может подходить больше за счет того, что она проходит через определенные точки

7 CAP

Сформулируем теоретически, что такое кривая CAP. Пусть у нас есть целевое распределение $\theta(\bar{x})$, которое мы хотим приблизить с помощью логистической регрессии $\theta^*(\bar{x})$. (Это функции от случайных величин, которые

возвращают значения от нуля до единицы). Определим функцию $\psi(t)$, которая возвращает множество $\psi(t) = \{\bar{x} \mid \theta^*(\bar{x}) \geq t\}$. Тогда ось x задается как

$$x(t) = P_{\theta^*}(\psi(t)) = 1 - F_{\theta^*}(t) \quad (9)$$

При этом

$$y(t) = \frac{\mathbb{E}_{\theta}[p(\bar{x}) \cdot \mathbb{I}(\bar{x} \in \psi(t))]}{\mathbb{E}_{\theta}[p(\bar{x})]} \quad (10)$$

Мы можем выразить $t(x)$, как

$$t(x) = F_{\theta^*}(1 - x)^{-1} \quad (11)$$

Следовательно

$$y(x) = \frac{\mathbb{E}_{\theta}[p(\bar{x}) \cdot \mathbb{I}(\theta^*(\bar{x}) \geq F_{\theta^*}(1 - x)^{-1})]}{\mathbb{E}_{\theta}[p(\bar{x})]} \quad (12)$$

для $x \in [0, 1]$

Если интерпретировать, данную формулу, то можно увидеть, что мы берем множество всех \bar{x} , которые входят в x самых редких предсказаний (т.е. не входят в $1 - x$ самых частых) и считаем, какой вклад в мат. ожидание делают эти \bar{x} в оригинальном распределении, т.е. считаем мат. ожидание учитывая только их (но не условное мат. ожидание). Форма идеальной кривой зависит сугубо от целевого распределения