Практическое задание 1: Байесовские рассуждения
Вариант 1

Антон Праздничных

5 Сентября 2021

Модель (1):

$$\begin{aligned} a &\sim \text{Unif}[a_{min}, a_{max}], \\ b &\sim \text{Unif}[b_{min}, b_{max}], \\ c|a,b &\sim \text{Bin}(a, p_1) + \text{Bin}(b, p_2), \\ d|c &\sim c + \text{Bin}(c, p_3) \end{aligned} \tag{1}$$

Модель (2) отличается от модели (1) лишь распределеинем $c|a,b = \text{Poiss}(ap_1 + bp_2)$

1. Далее нам нам потребуются следующие распеределения (помимо заданных в условии): $p(c)$, $p(d)$, $p(c|a)$, $p(c|b)$, $p(c|d)$, $p(c|a,b,d)$.

   Пойдем по порядку.

$$p(c) = \sum_{a,b} p(c|a,b)p(a)p(b) \tag{2}$$

   Для модели (1):

$$p(c=k|a,b) = \sum_{i=0}^{k} \text{Bin}(i|a,p_1)\text{Bin}(k-i|b,p_2) \tag{3}$$

   Для модели (2) это по условию распределеине Пуассона. Таким образом, мы знаем всерасределения, входящие в (2), так что формула корректна. Идем дальше

$$p(d) = \sum_{c} p(d|c)p(c) \tag{4}$$

   Найдем $p(d|c)$:

$$p(d=k|c) = \sum_{i=0}^{k} p(c=i)\text{Bin}(k-i|i,p3) \tag{5}$$

   Теперь (4) тоже корректно определено. Идем дальше

$$p(c|a) = \sum_{b} p(c|a,b)p(b) \tag{6}$$

   Тут уже все знаем. Аналогично

$$p(c|b) = \sum_{a} p(c|a,b)p(b) \tag{7}$$

   Наконец, пришло время воспользоваться теоремой Байеса:

$$p(c|d) = \frac{p(d|c)p(c)}{p(d)} \tag{8}$$

   И наконец

$$p(c|a,b,d) = \frac{p(d|c)p(c|a,b)}{p(d|a,b)p(a)p(b)} \tag{9}$$

2. Найдем матожидание и дисперсию случайных величин a, b, c d

(a) a и b распеределены равномерно на разных интервалах давайте для краткости выкладок найдем матожидание и дисперсию произвольной с.в. $\xi \sim U[a, b]$, а потом подставим нужные границы. Матожидание:

$$\mathbb{E}[\xi] = \frac{1}{b-a+1}\sum_{k=a}^{b} k = \frac{1}{b-a+1}\frac{(b+1)b - a(a-1)}{2} = \\ = \frac{b^2 - a^2 + b + a}{2(b-a+1)} = \frac{(b-a)(b+a) + (b+a)}{2(b-a+1)} = \frac{a+b}{2} \quad (10)$$

Второй момент:

$$\mathbb{E}[\xi^2] = \frac{1}{b-a+1}\sum_{k=a}^{b} k^2 = \frac{b(b+1)(2b+1) - (a-1)a(2a-1)}{6(b-a+1)} = \frac{2b^3 + 3b^2 + b - 2a^3 + 3a^2 - a}{6(b-a+1)} = \\ = \frac{(b-a)(2a^2 + 2ab + 2b^2) + (2a^2 + 2b^2 + 2ab) + (a^2 + b^2 - 2ab) + (b-a)}{6(b-a+1)} = \\ = \frac{(b-a+1)(2a^2 + 2ab + 2b^2) + (b-a)^2 + (b-a)}{6(b-a+1)} = \frac{(b-a+1)(2a^2 + 2ab + 2b^2) + (b-a)^2 + (b-a)}{6(b-a+1)} = \\ = \frac{2a^2 + 2ab + 2b^2 + b - a}{6} \quad (11)$$

Дисперсия:

$$\mathbb{D}\xi = \mathbb{E}\xi^2 - (\mathbb{E}\xi)^2 = \frac{2a^2 + 2ab + 2b^2 + b - a}{6} - \frac{a^2 + 2ab + b^2}{4} = \frac{4a^2 + 4ab + 4b^2 + 2(b-a) - 3a^2 - 6ab - 3b^2}{12} = \\ = \frac{(b-a)^2 + 2(b-a) + 1 - 1}{12} = \frac{(b-a+1)^2 - 1}{12} \quad (12)$$

Подставляя нужные чиселки, получаем:

$$\mathbb{E}a = \frac{a_{min} + a_{max}}{2} = \frac{90+75}{2} = 82.5; \quad \mathbb{D}a = \frac{(90-75+1)^2 - 1}{12} = \frac{255}{12} = \frac{85}{4} = 21.25 \quad (13)$$

$$\mathbb{E}b = 550; \quad \mathbb{D}b = \frac{10200}{12} = 850 \quad (14)$$

(b) Тут уже слишком жестокие формулы, чтобы считать аналитически, так что посчитаем численно. Получим следующие результаты:

	$\mathbb{E}a$	$\mathbb{E}b$	$\mathbb{E}c$	$\mathbb{E}d$
model 1	82.5	550	13.745	17.875
model 2	82.5	550	13.745	17.875

Таблица 1: Матожидания

	$\mathbb{D}a$	$\mathbb{D}b$	$\mathbb{D}c$	$\mathbb{D}d$
model 1	21.25	850	13.1675	25.140575
model 2	21.25	850	14.0475	26.627775

Таблица 2: Дисперсии

3. Посмотрим на то, как происходит уточнение прогноза для величины c по мере прихода новой косвенной информации для наших моделей (см 1, 3. 4).

   Видно, что наибольшую информацию о c несет в себе конкретное значение d: дисперсия таких распределений на порядок меньше.

4. Давайте проверим последнее утверждение для всех a, b, d а не только для их средних значений, а именно $\forall a, b, d\ \mathbb{D}[c|d] < \mathbb{D}[c|b] \wedge \mathbb{D}[c|d] < \mathbb{D}[c|a]$. Проверять, разумеется, будем численно. Честно говоря, при моей численной проверке получилось, что для $d \in [53, 64]$ ($d \in [53, 66]$ для модели 2) оба неравенство нарушаются, но, возможно, это артефакт численных расчетов, связанный с тем, что

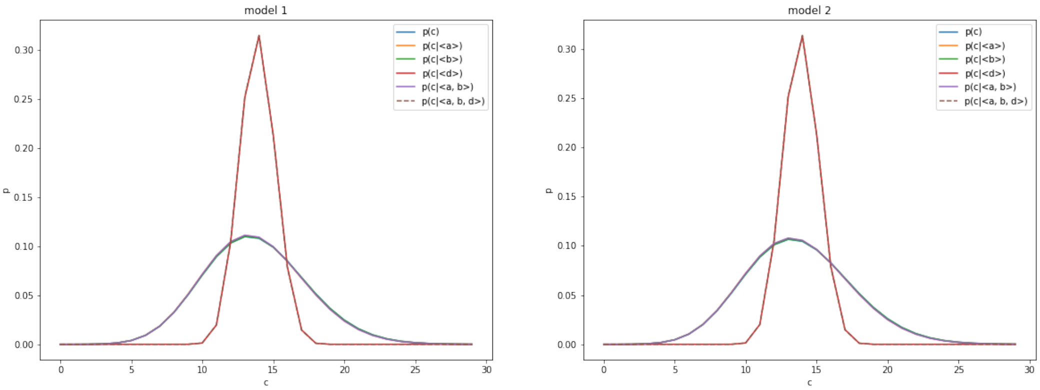

Рис. 1: Распределение c при различных данных

	$\mathbb{E}c$	$\mathbb{E}[c\|\bar{a}]$	$\mathbb{E}[c\|\bar{b}]$	$\mathbb{E}[c\|\bar{d}]$	$\mathbb{E}[c\|\bar{a},\bar{b}]$	$\mathbb{E}[c\|\bar{a},\bar{b},\bar{d}]$
model 1	13.745	13.7	13.75	13.896	13.7	13.891
model 2	13.745	13.7	13.75	13.894	13.7	13.889

Таблица 3: Матожидания $c|$.

	$\mathbb{D}c$	$\mathbb{D}[c\|\bar{a}]$	$\mathbb{D}[c\|\bar{b}]$	$\mathbb{D}[c\|\bar{d}]$	$\mathbb{D}[c\|\bar{a},\bar{b}]$	$\mathbb{D}[c\|\bar{a},\bar{b},\bar{d}]$
model 1	13.1675	12.91	13.0825	1.5336	12.825	1.5294
model 2	14.0475	13.785	13.9625	1.5439	13.7	1.5402

Таблица 4: Дисперсии $c|$.

для некоторых d в моем рассчете $p(c|d)$ происходило деление на 0, и я добавил регуляризацию (если что, реализация без регуляризации прошла все тесты ejudge).

Построив scatter plot множеств $S_1 = \{a, b | \mathbb{D}[c|b] < \mathbb{D}[c|a]\}$ и $S_2 = \{a, b | \mathbb{D}[c|b] \geq \mathbb{D}[c|a]\}$, легко видеть, что эти множества линейно разделимы для обоих модедей

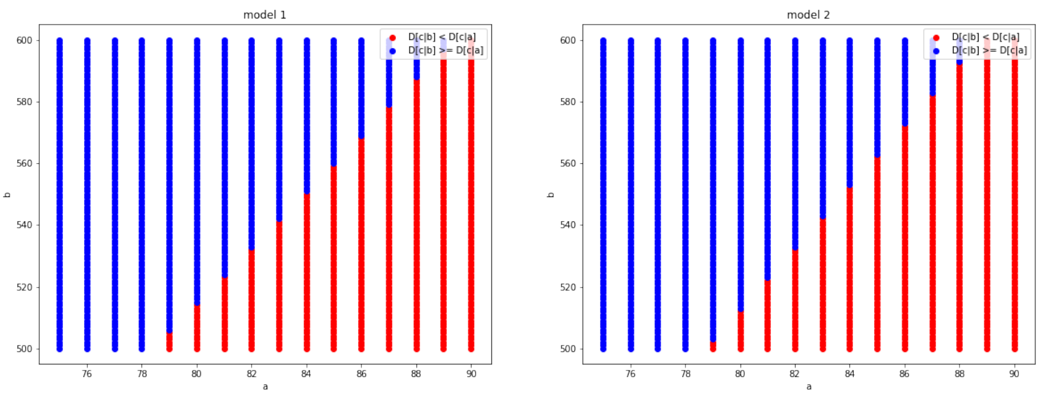

Рис. 2: Линейная разделимость S_1 и S_2

5. Замерим скорость вычислений, необходимых для оценки распределений $p(c)$, $p(c|a)$, $p(c|b)$, $p(c|d)$, $p(c|a,b)$, $p(c|a,b,d)$, $p(d)$ для скалярных a, b, d (равных своим средним значениям). Результаты приведены в таблице 5

   Видно, что почти везде упрощенная модель (2) кратно (в среднем раза в 3) выигрывает у модели (1).

6. На основе всех предыдущих пунктов можно сделать вывод, что модели (1) и (2) почти одинаковы и

	model 1	model 2
$p(c)$	555	158
$p(c\|a)$	53.9	10.1
$p(c\|b)$	15.2	3.53
$p(c\|d)$	627	201
$p(c\|a,b)$	8.37	2.03
$p(c\|a,b,d)$	93.4	89.6
$p(d)$	635	235

Таблица 5: Время оценки распределений, ms

отличаются главным образом временем оценки распределений. Очевидно, это связано с тем, что в упрощенной модели (2) для оценки распределения $c|a,b$ (и, следовательно всех $c|.$) необходимо вычислить лишь одну функцию плотности, в то время как в полной модели (1) необходимо выичслить 2 функции и еще сделать $\mathcal{O}(c_{max})$ операций для вычисления из них $p(c|a,b)$. Поскольку в остальном модели почти не отличаются, кажется, разумным для данной задачи использовать модель (2)