

## Наивная байесовская параметрическая модель

Наивная байесовская параметрическая модель использует простое предположение, чтобы резко уменьшить количество параметров для оценки. Мы *наивно* предполагаем, что каждый признак  $x_j$  *не зависит* от любого другого признака  $x_k$ .<sup>1</sup> В отношении набора данных, полученных на Ирме, это означает, например, что выбор цвета волос не влияет на выбор типа одежды, а выбор типа очков не влияет на выбор прически. Более формально, для всех признаков  $x_j, x_k$ :

$$p(x_j|x_k) = p(x_j).$$

Это известно как *наивное байесовское* предположение. Чтобы применить его, сначала используется цепное правило вероятности для записи функции плотности в виде произведения условных вероятностей:

$$\begin{aligned} p(\mathbf{x}) &= p(x_1, \dots, x_K) = \\ &= p(x_2, \dots, x_K | x_1) p(x_1) = \\ &= p(x_3, \dots, x_K | x_2, x_1) p(x_2 | x_1) p(x_1) = \\ &= \prod_{k=1}^K p(x_k | x_1, \dots, x_{k-1}), \end{aligned}$$

где  $K$  — общее число признаков (то есть пять в примере с планетой Ирм).

Теперь применим наивное байесовское предположение, чтобы упростить последнюю строку:

$$p(\mathbf{x}) = \prod_{k=1}^K p(x_k).$$

Это — наивная байесовская модель. Задача сводится к оценке параметров  $\theta_{kl} = p(x_k = l)$  для каждого признака в отдельности и их перемножению для определения вероятности любой возможной комбинации.

Сколько параметров нужно оценить в нашей задаче? Для каждого признака нужно оценить параметр для каждого значения, которое может принять этот

---

<sup>1</sup> Когда присутствует переменная отклика  $y$ , наивное байесовское предположение гласит, что существует *условная* независимость между каждой парой признаков  $x_j, x_k$  при заданном  $y$ .

признак. Следовательно, в примере с планетой Ирм эта модель определяется всего  $7 + 6 + 3 + 4 + 8 - 5 = 23$  параметрами.<sup>1</sup>

Оценка максимального правдоподобия  $\hat{\theta}_{kl}$  вычисляется как

$$\hat{\theta}_{kl} = \frac{n_{kl}}{N},$$

где  $\hat{\theta}_{kl}$  — число раз, когда признак  $k$  принимает значение  $l$  в наборе данных, а  $N = 50$  — общее число наблюдений.

В табл. 1.2 показаны вычисленные параметры для набора данных с планеты Ирм.

Чтобы найти вероятность, с которой модель сгенерирует некоторое наблюдение  $\mathbf{x}$ , достаточно перемножить вероятности отдельных признаков. Например:

$$\begin{aligned} & p(\text{длинная стрижка, прямые волосы; рыжий; круглые очки;} \\ & \quad \text{футболка с круглым вырезом; белый}) = \\ & = p(\text{длинная стрижка, прямые волосы}) \times p(\text{рыжий}) \times p(\text{круглые очки}) \times \\ & \quad \times p(\text{футболка с круглым вырезом}) \times p(\text{белый}) = \\ & \quad = 0,46 \times 0,16 \times 0,44 \times 0,38 \times 0,44 = \\ & \quad = 0,0054 \end{aligned}$$

Обратите внимание: эта комбинация отсутствует в исходном наборе данных, но наша модель определяет для нее ненулевую вероятность, а значит, вполне может сгенерировать ее. Кроме того, вероятность этой комбинации выше, чем, например, *(длинная стрижка, прямые волосы; рыжий; круглые очки; футболка с круглым вырезом; синий01)*, потому что белый цвет одежды появляется в наборе наблюдений чаще, чем синий.

То есть наивная байесовская модель способна выявить некоторую структуру данных и использовать ее для создания новых образцов, отсутствующих в исходном наборе. Модель оценила вероятность встретить каждое значение признака независимо от других, поэтому при использовании наивного байесовского предположения можно перемножить эти вероятности, чтобы построить полную функцию плотности,  $p_{\theta}(\mathbf{x})$ .

<sup>1</sup> Последний член в выражении  $-5$  отражает тот факт, что последний параметр для каждого признака подбирается так, чтобы сумма его параметров была равна 1.

**Таблица 1.2.** Оценки максимального правдоподобия для параметров в наивной байесовской модели

Прическа	n	$\hat{\theta}$	Цвет волос	n	$\hat{\theta}$	Цвет одежды	n	$\hat{\theta}$
нет волос	7	0,14	черный	7	0,14	черный	0	0,00
длинные волосы, собранные в пучок	0	0,00	блонд	6	0,12	синий01	4	0,08
длинная стрижка, волнистые волосы	1	0,02	каштановый	2	0,04	серый01	10	0,20
длинная стрижка, прямые волосы	23	0,46	пастельный розовый	3	0,06	пастельный зеленый	5	0,10
короткая стрижка, волнистые волосы	1	0,02	рыжий	8	0,16	пастельный оранжевый	2	0,04
короткая стрижка, прямые волосы	11	0,22	серебристо-серый	24	0,48	розовый	4	0,08
короткая стрижка, курчавые волосы	7	0,14	<i>Всего</i>	<i>50</i>	<i>1,00</i>	красный	3	0,06
<i>Всего</i>	<i>50</i>	<i>1,00</i>	<i>Всего</i>	<i>50</i>	<i>1,00</i>	белый	22	0,44
						<i>Всего</i>	<i>50</i>	<i>1,00</i>

Вид очков	n	$\hat{\theta}$
нет	11	0,22
круглые	22	0,44
солнцезащитные	17	0,34
<i>Всего</i>	<i>50</i>	<i>1,00</i>

Тип одежды	n	$\hat{\theta}$
Худи	7	0,14
комбинезон	18	0,36
футболка с круглым вырезом	19	0,38
футболка с V-образным вырезом	6	0,12
<i>Всего</i>	<i>50</i>	<i>1,00</i>

На рис. 1.8 показаны 10 наблюдений, выбранных моделью.

Для этой простой задачи наивное байесовское предположение о независимости признаков является разумным и, следовательно, дает хорошую генеративную модель.

Теперь посмотрим, что получится, если это предположение оказывается ошибочным.



**Рис. 1.8.** Десять новых стилей для ирмян, сгенерированных наивной байесовской моделью

## Привет, Ирм! Продолжение

Вы испытываете определенное чувство гордости, глядя на десять новых творений, созданных вашей наивной байесовской моделью. Воодушевленные своим успехом, вы обращаете внимание на другую сторону задачи, и на этот раз она не выглядит такой же простой.

Набор данных с незамысловатым названием Planet Pixel, который был вам предоставлен, не содержит пяти высокоуровневых признаков, которые вы видели выше (*цвет волос, тип аксессуара* и т. д.), а только значения  $32 \times 32$  пикселей, составляющих каждое изображение. То есть каждое наблюдение теперь имеет  $32 \times 32 = 1024$  признака и каждый признак может принимать любое из 256 значений (отдельные цвета в палитре).

Изображения из нового набора данных показаны на рис. 1.9, а выборка значений пикселей для первых десяти наблюдений показана в табл. 1.3.

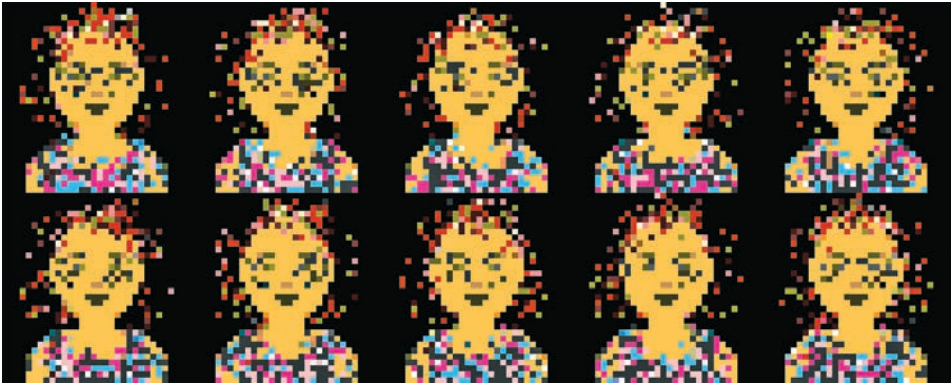
Вы решаете попробовать наивную байесовскую модель еще раз, на этот раз обученную на наборе данных пикселей. Модель оценит параметры максимального правдоподобия, определяющие распределение цвета каждого пикселя, чтобы на основе этого распределения сгенерировать новые наблюдения. Однако, закончив модель, вы понимаете, что что-то пошло не так. Вместо новых образцов моды модель вывела десять похожих друг на друга изображений, на которых нельзя различить ни аксессуары, ни четкие признаки прически или одежды (рис. 1.10). Почему так случилось?



Рис. 1.9. Образцы моды в Planet Pixel

Таблица 1.3. Значения пикселей 458–467 для первых десяти наблюдений в Planet Pixel

№ образца	px_458	px_459	px_460	px_461	px_462	px_463	px_464	px_465	px_466	px_467
0	49	14	14	19	7	5	5	12	19	14
1	43	10	10	17	9	3	3	18	17	10
2	37	12	12	14	11	4	4	6	14	12
3	54	9	9	14	10	4	4	16	14	9
4	2	2	5	2	4	4	4	4	2	5
5	44	15	15	21	14	3	3	4	21	15
6	12	9	2	31	16	3	3	16	31	2
7	36	9	9	13	11	4	4	12	13	9
8	54	11	11	16	10	4	4	19	16	11
9	49	17	17	19	12	6	6	22	19	17



**Рис. 1.10.** Десять новых стилей, сгенерированных наивной байесовской моделью на основе набора данных Planet Pixel

## Сложности генеративного моделирования

Во-первых, поскольку наивная байесовская модель выбирает пиксели независимо друг от друга, она не понимает, что два смежных пиксела обычно имеют похожие оттенки, потому что, например, являются частью одного и того же предмета одежды. Модель может сгенерировать цвет лица и рта, поскольку все эти пиксели в обучающем наборе имеют примерно одинаковый оттенок во всех наблюдениях; однако цвет пикселей для футболки выбирается случайно из множества цветов, присутствующих в обучающем наборе, независимо от цвета соседних пикселей. Кроме того, отсутствует механизм формирования пикселей в области вокруг глаз, позволяющий получить круглые очки, или в верхней части изображения, чтобы получить конкретную прическу.

Во-вторых, на этот раз в выборочном пространстве имеется непостижимо большое количество возможных наблюдений. Лишь небольшая часть из них — узнаваемые лица, а еще меньшее подмножество — лица, которые придерживаются правил моды в Planet Pixel. Следовательно, если наивная байесовская модель обучается на сильно коррелированных значениях пикселей, то вероятность найти удовлетворительную комбинацию значений оказывается очень мала.

В первом случае мы имели независимые признаки и относительно небольшое выборочное пространство, поэтому наивная байесовская модель показала неплохие результаты. Во втором случае, когда данные были представлены наборами пикселей, предположение о независимости значений пикселей оказалось несостоятельным. Значения пикселей сильно коррелированы, а выборочное пространство огромно, поэтому получить правильное изображение путем независимой выборки пикселей практически невозможно. Это объясняет, почему наивные байесовские модели плохо работают с обработанными изображениями.

Рассмотренный нами пример высветил две ключевые проблемы, которые должна преодолеть генеративная модель, чтобы добиться успеха.

#### **СЛОЖНОСТИ ГЕНЕРАТИВНОГО МОДЕЛИРОВАНИЯ**

- Как модель может справиться с высокой условной взаимозависимостью признаков?
- Как модель может отыскать одну из крошечных пропорций, чтобы получить удовлетворительное наблюдение в многомерном выборочном пространстве?

Ключом к решению обеих этих проблем является глубокое обучение.

Нам нужна модель, способная выявить релевантную структуру в данных, которая не требует делать каких-либо предположений заранее. Именно с этим прекрасно справляется глубокое обучение, и именно поэтому этот метод стал движущей силой последних достижений в генеративном моделировании.

Тот факт, что глубокое обучение может формировать свои признаки в пространстве более низкой размерности, означает, что это — форма *обучения представлению*. Прежде чем приступить к глубокому обучению в следующей главе, мы должны разобраться с ключевыми понятиями обучения представлению.