Capítulo 3
Dominar los conceptos básicos

1. Estar en armonía con los datos

El error más común que cometen las personas recién iniciadas en data sciences es usar directamente, sin precauciones y sin un análisis preliminar, los algoritmos disponibles sobre datos mal entendidos.

A lo largo del proceso de desarrollo de su estudio y desde su primer contacto con un problema, el data scientist estará interesado en evaluar sus datos, usando diferentes herramientas estadísticas para poder visualizarlos desde varios puntos de vista. Esta es la única manera de poder formular con juicio las hipótesis que queremos validar, o no, a lo largo del proceso.

Observación

El lector interesado que quiera volver a los aspectos básicos, sin duda quedará satisfecho con el libro de Andrei Kolmogorov, a quien debemos muchas contribuciones a ciertos conceptos que subyacen a nuestra práctica de las data sciences (formalización de la lógica intuitiva como «cálculo sobre problemas», ley fuerte de los grandes números, axiomas de probabilidad a través del lenguaje de las teorías de medida, upsilon-entropía que permite calificar los estimadores, aplicación del método de Poincaré al estudio de equilibrios, etc.).

La interpretación de herramientas estadísticas requiere asimilar algunas nociones fundacionales sobre probabilidad, que vamos a abordar ahora.

1.1 Algunas nociones principales

1.1.1 Fenómeno aleatorio

Se enfrenta a un **fenómeno aleatorio** si, cuando se realiza un **experimento** varias veces en condiciones idénticas, se obtienen **resultados** diferentes (y, por lo tanto, en cierto modo, impredecibles).

Observación

Observe que, en esta definición, el hecho de que un fenómeno se identifique como aleatorio depende de la incapacidad del observador para preverlo por completo, lo que le hará declarar de forma categórica que el fenómeno es intrínsecamente aleatorio.

Sin embargo, cuando repetimos el experimento un gran número de veces, vemos que los resultados se distribuyen siguiendo una ley o distribución estables, es decir, **con frecuencias de aparición de los resultados que dependen del valor del resultado**. Por ejemplo, si observa las calificaciones de bachillerato de un gran número de estudiantes, encontrará una frecuencia de las calificaciones promedio más altas que la frecuencia de las calificaciones peores o buenas. Debido a que tenemos esta **ley de los grandes números**, podemos permitirnos hacer algunos cálculos matemáticos para estudiar el fenómeno.

Un **evento, que se define como una parte de todos los resultados posibles**, es tanto más probable cuanto más grande es la parte de los resultados posibles.

Por ejemplo, el evento **`tener_una_nota_superior_a_5`** tiene una probabilidad mayor que el evento **`tener_una_nota_superior_a_8`**.

Observación

*Cuando el número de posibles resultados es infinito, el número de posibles eventos también lo es y surgen ciertas dificultades matemáticas. En este caso, los matemáticos trabajan en un subconjunto del conjunto de eventos posibles, al que llaman **tribu**. Una tribu es estable mediante intersección y unión numerable de eventos (que se pueden contar), es decir, estas intersecciones o uniones pertenecen a la tribu. Dado un evento de la tribu, esta incluye como parte de la tribu los eventos vacío y su opuesto o complementario, es decir, «todos los resultados posibles». En este caso, solo podremos definir una probabilidad sobre los miembros (eventos) de la tribu.*

1.1.2 Probabilidad, variable aleatoria y distribución

Probabilidad

Una **probabilidad es una aplicación** que mapea los eventos de una tribu, con valores entre 0 y 1 (ambos incluidos), tal que:

- la probabilidad del evento, incluidos todos los resultados posibles, es igual a 1,
- la probabilidad de cualquier unión numerable de eventos disjuntos de la tribu (es decir, sin un resultado común) es igual a la suma de las probabilidades de estos eventos.

Por ejemplo, la probabilidad de tener una nota entre 0 y 20 es igual a 1, y la probabilidad de tener una nota entre 16 y 18 **o** entre 18 y 20 es igual a la probabilidad de tener una nota entre 16 y 20.

Además, la probabilidad de un evento imposible (por ejemplo, tener una puntuación inferior a 5 y al mismo tiempo superior a 18) obviamente es cero.

Variable aleatoria y distribución

Veamos ahora la noción de variable aleatoria porque puede generar confusión. En efecto, **una variable aleatoria no es una variable, sino una función**, que va desde los resultados de un experimento de un espacio medible hasta un conjunto no especificado y que nos permite caracterizar el experimento.

Supongamos que está interesado en los libros de una biblioteca y la variable aleatoria **x** representa el número de páginas. Existe una función que, para cada resultado (es decir, cada libro elegido al azar), devuelve el número de páginas. Evidentemente, son posibles otras variables aleatorias; por ejemplo, una variable **y** que en nuestro ejemplo indicará si la cubierta es dura o blanda.

Seguidamente, obtenemos una probabilidad $\mathbf{p}_x$, que se llama **ley de la variable x** (o **distribución de x** si nos referimos a la estadística), que proporciona una probabilidad para cada valor de **x** (aquí, el número de páginas) y para cada evento de la tribu de los posibles subconjuntos de los valores de **x** (por ejemplo, la probabilidad de elegir una obra de entre 100 y 200 páginas). Podemos ver fácilmente que el aspecto de la distribución de **x** no tiene nada que ver con el de la distribución de **y** (tapa dura o blanda).

Cálculos sencillos sobre las distribuciones

Hay algunas propiedades que debe conocer sobre las distribuciones.

Tenemos: $\mathbf{p}_{x \text{ e } y} + \mathbf{p}_{x \text{ o } y} = \mathbf{p}_x + \mathbf{p}_y$

Si **x** e **y** son disjuntos, es decir, mutuamente excluyentes, tenemos: $\mathbf{p}_{x \text{ e } y} = \mathbf{0}$ y por tanto $\mathbf{p}_{x \text{ o } y} = \mathbf{p}_x + \mathbf{p}_y$.

Estas últimas líneas se entienden fácilmente si tomamos un punto de vista conjuntista donde la expresión **x e y** significa **la intersección** de los eventos correspondientes, y donde la expresión **x o y** significa la **unión** de los eventos correspondientes.

El ejemplo de la siguiente sección describe una forma de determinar $\mathbf{p}_{x\ e\ y}$ y le ayudará a entender mejor esta noción.

Condicionamiento y probabilidad condicional

Cuando hemos fijado el valor o valores de una variable aleatoria cuya distribución conocemos y si las variables son dependientes, esta información condiciona la probabilidad de otra variable.

Para simplificar las siguientes explicaciones, vamos a discretizar la variable aleatoria **x**, es decir, la vamos a transformar en un número finito de valores. Consideraremos que un libro de menos de 100 páginas es «pequeño» y «grande» en caso contrario.

Consideremos en general que el 90 % de los libros «grandes» utilizan tapas duras y, en el caso contrario, dichas tapas solo se utilizan en el 15 % de los libros. Ahora tenemos una distribución de **probabilidad condicional** que podemos denotar por $\mathbf{p}_{y|x}$ y que se establece de la siguiente manera: **`p_de_y_sabiendo_x`**.

Veamos en detalle el aspecto de esta distribución de probabilidad condicional. Lo expresamos diciendo «la probabilidad de que el libro tenga tapa dura **sabiendo que es grande** es del 90 %»:

```
p_y|x(dura   , libro_grande  ) = 90 %  # entonces:
p_y|x(blanda , libro_grande  ) = 10 %
p_y|x(dura   , libro_pequeño ) = 15 %
p_y|x(blanda , libro_pequeño ) = 85 %
```

Si conocemos la distribución $\mathbf{p}_x$ de los libros grandes y pequeños de la biblioteca, podemos inferir la distribución de cubiertas duras gracias al tamaño de los libros. Supongamos que tenemos un 80 % de libros grandes, entonces:

```
p_x(libro_grande  ) = 80 %     # y:
p_x(libro_pequeño ) = 20 %
```

Esto nos permite calcular la distribución de cada tipo de libro con respecto a las dos variables aleatorias juntas y hacerlo sobre todos los libros de la biblioteca: $\mathbf{p}_{y\ y\ x}$. La probabilidad de que un libro sea de tapa dura y grande:

```
p_y y x(dura,libro_grande)=
                  p_y|x(dura,libro_grande) * p_x(libro_grande) =
                  90% * 80% = 72% = 0.72

# entonces:

p_y y x(blanda,libro_grande)=
```

$\mathbf{p}_{y|x}$(blanda ,libro_grande) * $\mathbf{p}_x$(libro_grande) =
10% * 80% = 8 % = 0.08

$\mathbf{p}_{y\ y\ x}$(dura,libro_pequeño)=
$\mathbf{p}_{y|x}$(dura,libro_pequeño) * $\mathbf{p}_x$(libro_pequeño) =
15% * 20% = 3% = 0.03

$\mathbf{p}_{y\ y\ x}$(blanda,libro_pequeño)=
$\mathbf{p}_{y|x}$(blanda,libro_pequeño) * $\mathbf{p}_x$(libro_pequeño) =
85% * 20% = 17% = 0.17

Observación

Vemos que la suma de $\mathbf{p}_{y\ y\ x}$ es igual a 1, que efectivamente se corresponde con el valor de la probabilidad cuando consideramos un evento que incluye todos los resultados posibles.

Se demuestra que podemos generalizar de la siguiente manera para variables aleatorias discretas (numerables), así como para distribuciones sobre variables aleatorias que reciben sus valores de conjuntos no numerables, como los números reales o los vectores reales (o incluso funciones):

$$\mathbf{p}_{y\ e\ x} = \mathbf{p}_{y|x} \cdot \mathbf{p}_x = \mathbf{p}_{x|y} \cdot \mathbf{p}_y = \mathbf{p}_{x\ e\ y}$$

Esta fórmula es válida asumiendo que las distribuciones de **x** e **y** no son nulas.

En la misma línea, con tres variables aleatorias tenemos:

$$\mathbf{p}_{x\ e\ y\ e\ z} = \mathbf{p}_x \cdot \mathbf{p}_{y|x} \cdot \mathbf{p}_{z|(x\ e\ y)}$$

Y así sucesivamente. Podemos generalizar para cualquier número de variables aleatorias.

Independencia

La noción de independencia entre variables es muy útil. En muchas ocasiones, es la condición estricta para poder utilizar ciertos algoritmos. De manera intuitiva, pensamos que es más fácil interpretar un fenómeno cuando podemos vincularlo a variables aleatorias independientes. La idea general es que tener información sobre una de las variables independientes no proporciona información sobre la distribución de la otra variable aleatoria.

Obviamente, esto se traduce en (asumiendo distribuciones distintas no nulas):

$\mathbf{p}_{x|y} = \mathbf{p}_x$ lo que es equivalente a:

$\mathbf{p}_{y|x} = \mathbf{p}_y$ lo que es equivalente a:

$\mathbf{p}_{x\ e\ y} = \mathbf{p}_x \cdot \mathbf{p}_y$

Observación

Tenga en cuenta que esta noción de independencia no se debe confundir con la noción de eventos disjuntos, es decir, que no comparten ningún valor de resultado. Por ejemplo, imagine que, en nuestra biblioteca, los colores del reverso de los libros dependen de la nacionalidad de los autores, que solo pueden ser franceses o alemanes. El evento {azul, blanco, rojo} significa que un libro tiene, al menos, un autor español y el evento {negro, rojo, amarillo} significa que tiene, al menos, un autor alemán. Estos dos eventos tienen una intersección no nula, a saber, el color {rojo}. Sin embargo, el hecho de identificar que un autor es alemán a través de los colores {negro, rojo, amarillo} no proporciona ninguna información sobre si existe o no otro autor del libro que sea francés y que se puede identificar por sus colores {azul, blanco, rojo} y viceversa. Las dos variables aleatorias son independientes y, sin embargo, no disjuntas.

1.1.3 Un poco de matemáticas: notación y definiciones útiles

Desde el inicio del capítulo, hemos ahorrado en notaciones para no sobrecargar la exposición. A continuación, vamos a ver las notaciones y definiciones a las que nos podríamos enfrentar.

Lista de notaciones muy básicas

La definición de un conjunto utiliza llaves. El conjunto A de números enteros pares se podría escribir como:

$A = \{\text{los enteros pares}\}$

$= \{n \in \mathbb{N} \mid n \text{ par}\}$ # la barra se lee «tal que»

$= \{0, 2, 4, \ldots\}$

$= \{n \in \mathbb{N} \mid \exists\, k \in \mathbb{N}, (n = 2k)\}$

Se puede leer «conjunto de los elementos del conjunto de enteros naturales, que aquí llamaremos n, tales que existe un número natural, que aquí llamaremos k, de tal manera que n es igual a 2 veces k».

Para denotar un experimento aleatorio, normalmente se usa el símbolo ε.

El espacio de estados asociado con el experimento (es decir, el conjunto de todos los resultados posibles del experimento aleatorio ε) se denota con omega mayúscula: Ω.

Uno de los resultados de este conjunto se denota (omega en minúsculas): ω.

El hecho de que este resultado pertenece a Ω se denota con $\omega \in \Omega$.

Si Ω es numerable, podemos escribirlo como el conjunto discreto de resultados posibles $\{\omega_1, \omega_2 \dots\}$ (aquí hemos indexado de 1 a infinito: $\mathbb{N}^*$): $\Omega = \{\omega_i\}_{i \in \mathbb{N}^*}$.

Si Ω es finito con una cardinalidad (es decir, un tamaño) de n elementos, podemos escribirlo como el conjunto discreto de resultados posibles: $\{\omega_1, \omega_2 \dots, \omega_n\}$: $\Omega = \{\omega_i\}_{i \le n, i \in \mathbb{N}^*}$.

El conjunto de partes de Ω, es decir, todos los conjuntos que podemos construir con elementos de Ω, al que agregamos el conjunto vacío ϕ, normalmente se escribe $\mathcal{P}(\Omega)$.

Notaciones y elementos relacionados con la teoría de las medidas

Detrás del término «medida» pensamos de manera intuitiva que estamos frente a una teoría que nos permitirá controlar la forma en que podemos percibir las cantidades (números reales positivos o cero), que se corresponderán con conjuntos de «alguna cosa».

Debemos ser conscientes de que no todo es medible y, por lo tanto, se ha impuesto a los matemáticos la necesidad de diseñar una teoría que permita una medición precisa. Los siguientes elementos introducen el vocabulario que permite leer textos que manejan esta noción de medida. No pensamos que fuera razonable evitar estos conceptos, un poco abstractos, en un libro sobre data sciences porque los data scientists manipulan probabilidades que son medidas, en el sentido matemático del término.

Vamos a empezar definiendo un término extraño en este contexto: «tribu». Una tribu $\mathcal{A}$ sobre Ω es un conjunto de subconjuntos resultantes de las partes de Ω : $\mathcal{P}(\Omega)$, que incluye el conjunto vacío ϕ, que es estable por complementariedad (el complementario en Ω de cualquier parte también pertenece a Ω) y también es estable por unión numerable (cualquier unión finita de conjuntos pertenecientes a la tribu también pertenece a la tribu). Como el complemento del conjunto vacío ϕ es Ω mismo, de acuerdo con la regla anterior, cualquier tribu incluye al menos los dos elementos de la siguiente tribu: $\{\phi, \Omega\}$.

Es útil saber que la unión o intersección finita de tribus Ω es, en sí misma, una tribu de Ω.