

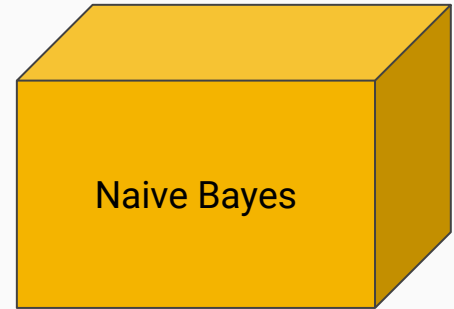
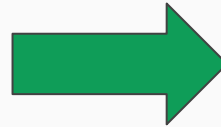
DeepSets

¿Cómo entrenamos un modelo?

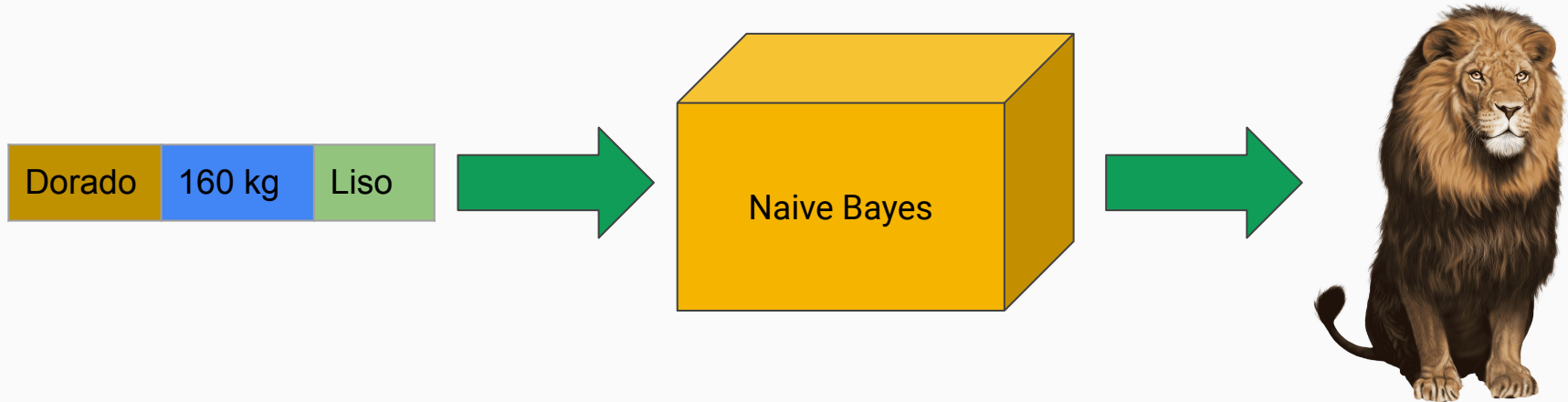
Por ejemplo, con Naive Bayes dadas características de felinos podemos predecir qué felinos son.

Color de pelo	Peso	Patrón del pelaje	Felino
Negro	70 kg	Liso	Pantera
Naranja	180 kg	Rayado	Tigre
⋮	⋮	⋮	⋮
Blanco y negro	5 kg	Liso	Gato siamés

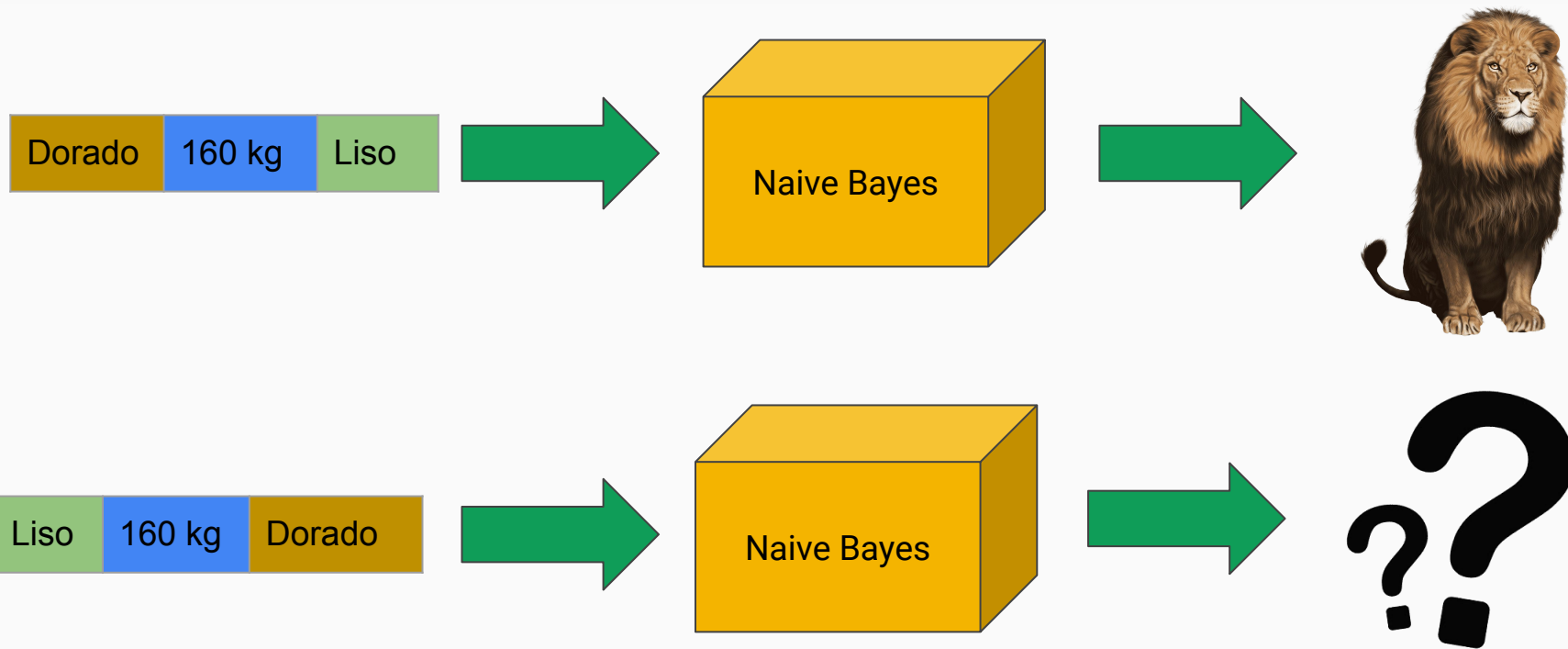
Se entrena
el modelo



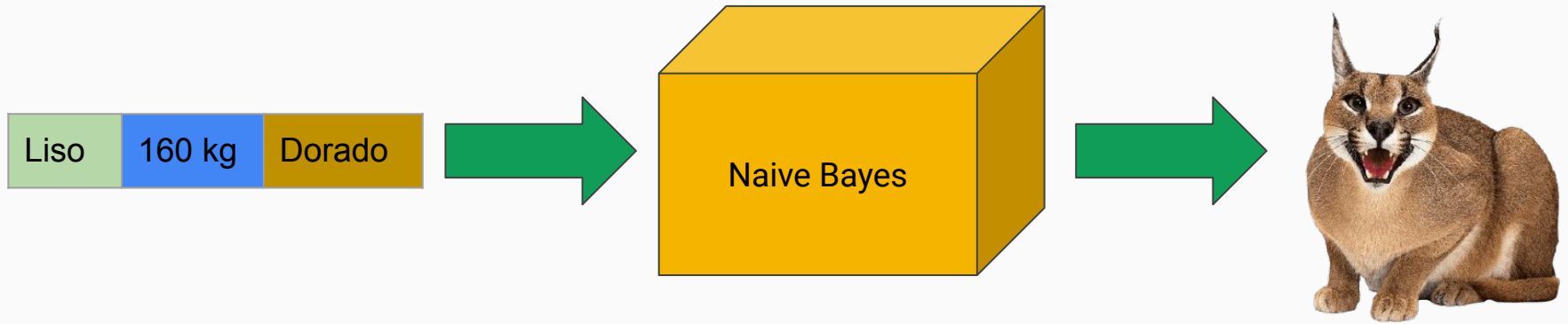
¿Cómo predecimos una nueva entrada?



¿Qué diferencias de predicción tendrán esta entrada en las paqueterías de machine learning?



El resultado esperado podría ser erróneo.



El problema:

Implícitamente al entrenar el modelo damos por input arreglos, los cuales dan importancia al orden.

Esto no sucedería si pensamos a los inputs como conjuntos.

En este caso:

Color

Peso

Patrón del pelaje

Deep Sets

Artículo:

<https://arxiv.org/pdf/1703.0611>

4

Autores:

- Manzil Zaheer
- Stawik Kottur
- Siamak Ravanbakhsh
- Barnabás Póczos
- Ruslan Salakhutdinov
- Alexander J Smola

Características de Deep Sets.

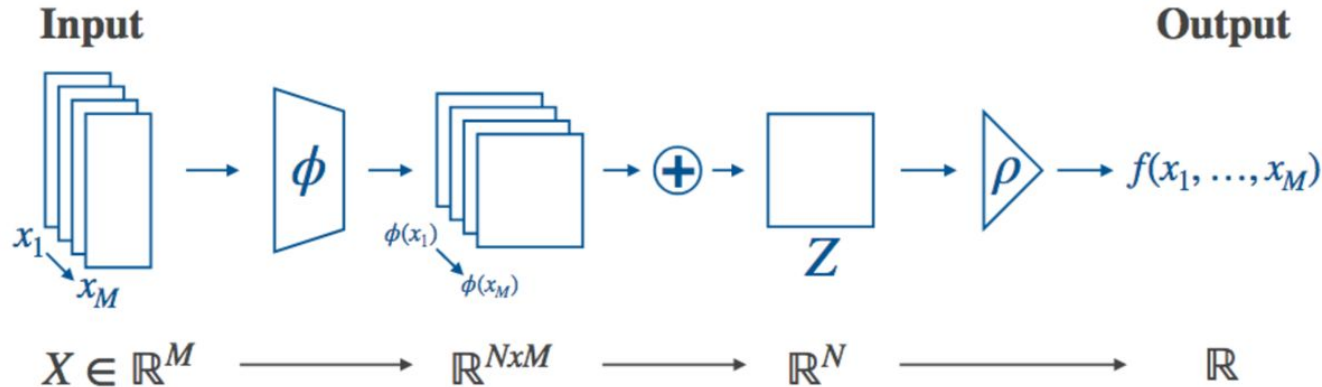
- Deep Sets es una red neuronal que tiene como objetivo que dado un input, el output debe de ser independiente del orden de las características del input. Es decir trata a los inputs como conjuntos.
- No es necesario entrenar para cada una de las permutaciones del conjunto para que se aprenda la invarianza.
- ¿Cómo podrían representar los conjuntos $\{1,2,3\}$ y $\{2,3,1\}$ con un número para que no importe el orden?

Funcionamiento:

- Teorema: Consideremos X un conjunto numerable y $f: X \rightarrow Y$. La función f es invariante bajo permutaciones de los elementos de X si y sólo si, $f(x) = \rho(\sum_{x \in X} \phi(x))$ para algunas funciones ρ y ϕ .
- Lema: Del teorema anterior, consideramos f una conexión de dos capas de una red neuronal de R^K a R^M , tenemos $f(x) = \sigma(\theta x)$ donde θ es la matriz de pesos y σ es de activación. Tenemos que la red neuronal es equivariante por las permutaciones si y sólo si los elementos de la diagonal θ son iguales entre sí y los que no también, pero no necesariamente iguales a los de la diagonal. Esto se reduce a $\theta = \alpha \otimes \text{Id} + \beta \otimes J + \gamma \otimes [1, \dots, 1]$ con λ, γ matrices arbitrarias de $M \times K$, Id la identidad de $K \times K$.

Arquitectura:

- Las capas de input se proyectan en un espacio de dimensión superior para tratar de identificar de manera única a cada elemento del conjunto.
- Se suman sobre las proyecciones y posteriormente se hace la sumatoria y se predice.
- Este proceso a veces se repite una y otra vez, para después llegar al output.



Aplicaciones:

- Suma de dígitos.
- Clasificación de nubes de puntos.
- Expansión de conjuntos de acuerdo a sus características.
- Etiquetar imágenes.