

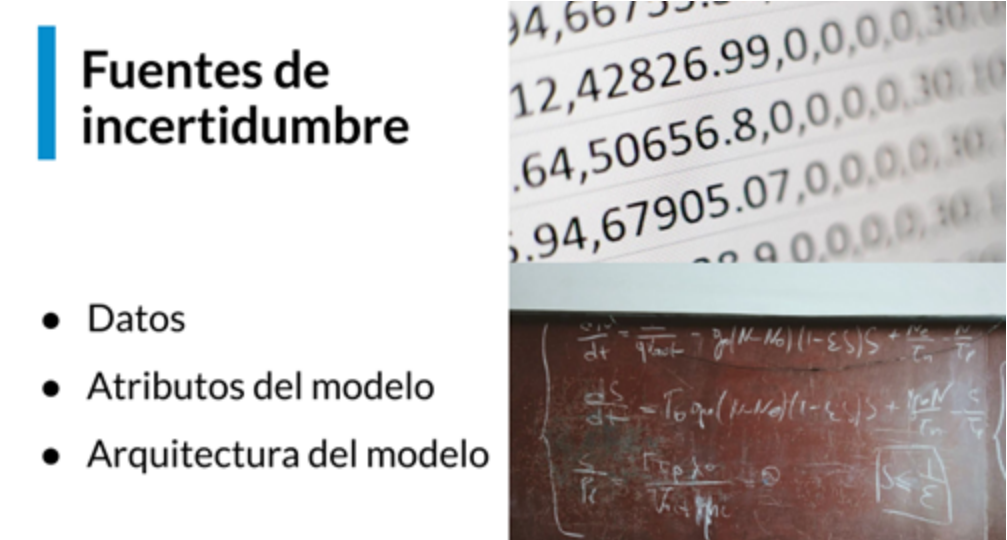
Probabilidad en Machine Learning

Veamos con en Machine Learning se usan las herramientas de probabilidad.

Recordemos que la probabilidad es un conjunto de herramientas que nos permiten cuantificar la incertidumbre y en ML ¿donde esta la incertidumbre? ¿cuál es la fuente de incertidumbre?

Hablemos de las principales 3:

- Datos
- Atributos del modelo
- Arquitectura del modelo



Datos

¿Por qué los datos? Recuerda que en la vida real, recoger y hacer medición de datos es un proceso imperfecto, todos los instrumentos de medición tienen un margen de error y este margen de error ya está introduciendo parte de incertidumbre en nuestros datos, y por lo tanto representa información imperfecta e incompleta.

Atributos del modelo

En Machine Learning siempre hablamos que un modelo se alimenta con atributos o variables que también llamamos predictores, estas variables con frecuencia son un sub-conjunto reducido de toda la realidad del problema que yo estoy intentando entender o sobre el cual quiero aplicar un modelo. Esta reducción de variables ya representa también otra capa de incertidumbre.

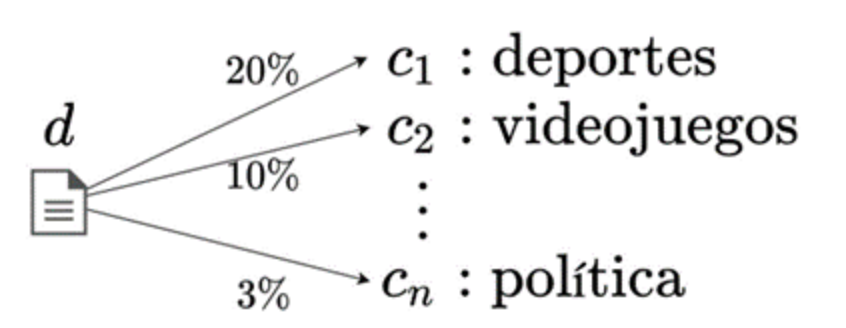
Modelos

En matemáticas los modelos siempre se entiende como una representación simplificada de la realidad y al ser esto así, esa representación simplificada por construcción induce otra capa de incertidumbre, ya que es información incompleta. Es menos información que yo estoy considerando, solo por el hecho de simplificar la realidad con mi modelo.

Así pues tenemos 3 grandes fuentes de incertidumbre y estas 3 fuentes de incertidumbre son cuantificables con elementos de probabilidad.

Modelos de clasificación

Veamos esto con un ejemplo particular **un modelo de clasificación** en el caso particular de documentos de texto.

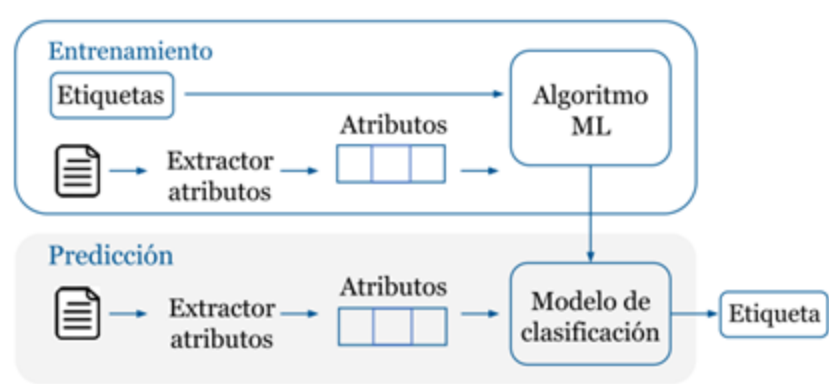


Tenemos un conjunto de documentos que cada uno tiene un tema particular al cual se refiere. Podemos tener documentos que hablen de; deportes, videojuegos, en fin n categorías donde la ultima puede ser "política" y el modelo tiene que leer cada uno de los documentos y con base en ello, ser capaz de identificar cual es el tema de conversación o cual es la categoría a la cual el documento pertenece. Y los modelos probabilísticos que hacen clasificación de este estilo funcionan precisamente asignando una probabilidad a cada una de las categorías digamos referenciando este documento, donde entonces la probabilidad mas alta es la probabilidad mas alta que me va indicar a que temática pertenece el documento. De pronto tiene algo de temática de videojuegos, pero la probabilidad mas alta es la de deportes y el modelo digamos que arroja como resultado que la categoría predicha es realmente de deportes.

Esto es lo que llamamos un clasificador probabilístico y aquí por definición, por construcción del modelo ya esta la probabilidad, ya que estamos reduciendo el problema precisamente al uso de estos números, para identificar cual es esa categoría más probable.

Veamos un poco más en detalle como es el funcionamiento interno de un modelo.

Funcionamiento



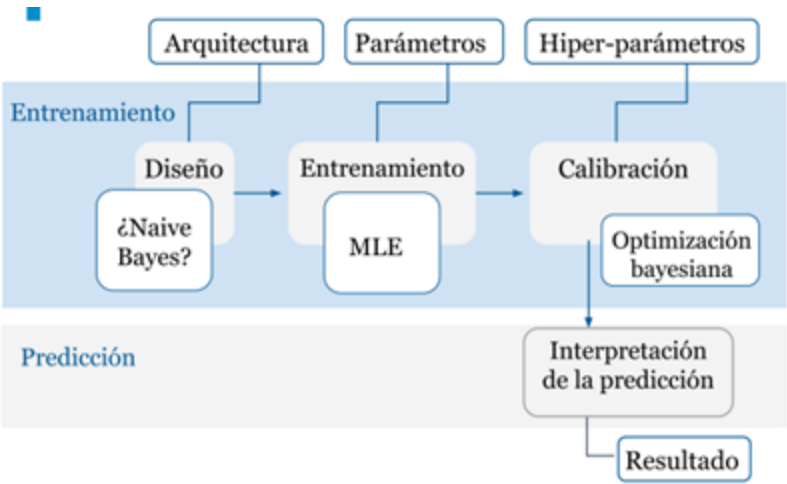
El modelo de clasificación de nuestro caso lo que hace es:

1. Tiene unos documentos que cada uno de ellos tiene etiqueta (La etiqueta es la categoría del tema al cual se refiere).
2. Los documentos pasan por una función que extrae los atributos; es decir simplifica las variables, el modelo no recoge todo el texto, sino que tu construyes el modelo entendiendo cuales son los elementos fundamentales de ese documento que van a hacer la predicción de la categoría y eso es lo que llamamos el **extractor de atributos**.
3. Una vez que hago eso, el documento se reduce a un conjunto de atributos que puede ser un vector, y ese vector se pasa como **INPUT(entrada)** al modelo; es decir al **algoritmo de Machine Learning de clasificación** y como este algoritmo lo estamos entendiendo lo llamamos **Supervisado**, porque también le estoy dando las etiquetas; es decir el algoritmo lee los atributos y con base en ello, el sabe cual debería ser el resultado es decir la etiqueta. Y así es como entrenamos en Machine Learning supervisado.
4. Luego a toda la fase anterior de entrenamiento le sigue la fase de **predicción**, en el proceso en el cual aprendió a juntar los atributos con las etiquetas correctas, después tomamos el modelo y le pasamos otros documentos. Estos documentos pasan por el mismo proceso de extracción de atributos y luego entran en el modelo de clasificación y el modelo predice la etiqueta, entonces yo puedo en general pasar documentos que no están etiquetados porque yo simplemente no se de que hablan.

Digamos tengo una tarea en la que tengo que leer muchos documentos que son muy extensos, y tomamos en cuenta que mi algoritmo está bien entrenado, yo le paso el modelo a mis documentos, entonces el algoritmo me va a dar la clasificación de los documentos. Y este es el procesamiento digamos completo de este modelo de clasificación y en general de todos los modelos de clasificación, todos funcionan de esa manera, con ese esquema general.

Así todas las etapas de un modelo en ciertos aspectos involucran probabilidad ¿cómo?

Etapas del modelo que involucran probabilidad



En la parte del entrenamiento, lo primero antes de pasarle los documentos para extraer atributos, identificar la arquitectura. Es precisamente escoger el modelo que voy a usar, al escoger el modelo que voy a usar, entonces es lo que llamamos el diseño. Y en el diseño es que vamos a decidir si el modelo va a usar probabilidad o no? **OJO AQUÍ, PORQUE NO TODOS LOS MODELOS USAN PROBABILIDAD**, no todos los modelos son probabilísticos.

Aquí ponemos un ejemplo de modelo **NAIVE BAYES** que es un modelo que vamos a ver mas adelante en el curso, entonces yo elijo este modelo precisamente porque es probabilístico y funciona bastante bien para esta tarea, luego de que yo escojo el diseño.

Tengo que definir el entrenamiento, y el entrenamiento es básicamente que el modelo aprenda la distribución de probabilidad. Y es una manera que yo uso para saber que probabilidades le asigno a cada una de las posibles ocurrencias de mis datos. Y entonces ahí hay un esquema que se llama **"MLE** Estimación de Maxima Verosimilitud o Maximum Likelihood Estimation en Inglés, que es una técnica probabilística.

Luego de eso, esta la calibración. Y es el ajuste de mi modelo, entonces el modelo posee algo que se llaman **parámetros** y que se calibraron en el entrenamiento, pero en lo que nos referimos a calibración como tal, son esos hiperparametros, que son en realidad parámetros que yo no puedo configurar con la optimización, (Esto se ve más a fondo en redes neuronales donde digamos el numero de neuronas en la capa, tiene digamos 10 neuronas y cada neurona tiene unos pesos que conectan las neuronas y esos pesos tu los vas calibrando de manera que el error del modelo sea cada vez mas pequeño) sin embargo hay parámetros de los modelos que tu no puedes calibrar a medida que buscas que el error disminuya, esos parámetros que están por fuera del proceso de optimización del modelo los llamamos **hiperparametros**, porque están por fuera de todo ese esquema de optimización. Y entonces el rectángulo de la ilustración anterior que dice **Calibración** a veces se le llama *Tuneo o calibración de hiperparametros*, ahí a veces también se usan algoritmos probabilísticos como la *optimización Bayesiana*.

Entonces es muy importante tener en cuenta que la arquitectura determina si es probabilístico o no, los parámetros como se van a entrenar para que el error del modelo sea el mínimo posible y los hiperparametros es; de todas las posibles arquitecturas del modelo, porque cada modelo en si

mismo es una familia donde yo puedo calibrar agregando o quitando mas peso en algo, determina también ese otro conjunto de hiperparametros que es muy importante para escoger cual de todas las familias de modelos funciona mejor.

Luego del proceso del entrenamiento del modelo viene una fase que se llama **predicción** y dentro de ella hay una parte que entendemos como **Interpretación del modelo**; aveces cuando un modelo de Machine Learning es demasiado complejo, se hace difícil su interpretación y con frecuencia si su modelo es probabilístico, requiere que la interpretación también use conceptos de probabilidad, la persona que esta interpretando el modelo debe entender muy bien como funcionan los cálculos de probabilidades del modelo, para obtener una interpretación correcta del mismo.

Y asi es como vemos en diferentes partes de todo el procesamiento del modelo, el entrenamiento y la predicción como es que se usa la Probabilidad. Esto es en general, no solo para modelos de clasificación sino para muchas otras tareas de Machine Learning Supervisado y no Supervisado.

En la proxima clase veremos que hay diferentes tipos de probabilidad y como se utilizan en diferentes partes de Data Science y Machine Learning.