



Curso de Introducción al pensamiento probabilista.

Introducción a la programación probabilista:

En la programación estocástica introducimos la aleatoriedad y las simulaciones para poder trabajar con la probabilidad.

PYRO \Rightarrow uber

Existen librerías específicas.

Uber necesita unir a los conductores con los clientes.

Tiene una gran cantidad y generar una hipótesis y cambiarlos.

Filtros de spam \Rightarrow aplicaciones de machine learning iniciando con programación probabilista.

Se junta evidencia.

La palabra y reduce la probabilidad de que sea verdad una hipótesis debido a que tienen que cumplirse más opciones.

Probable que ambas cosas sucedan.

Probabilidad Condicional

Probabilidad \Rightarrow Cuantas veces puede suceder un evento dentro de todas las posibilidades.

Probabilidad independiente \Rightarrow los eventos no esta relacionadas con los próximos.

Probabilidad condicional \Rightarrow tomar en consideración el evento anterior. en base a que se cumpla la condición en conjunto con otro.

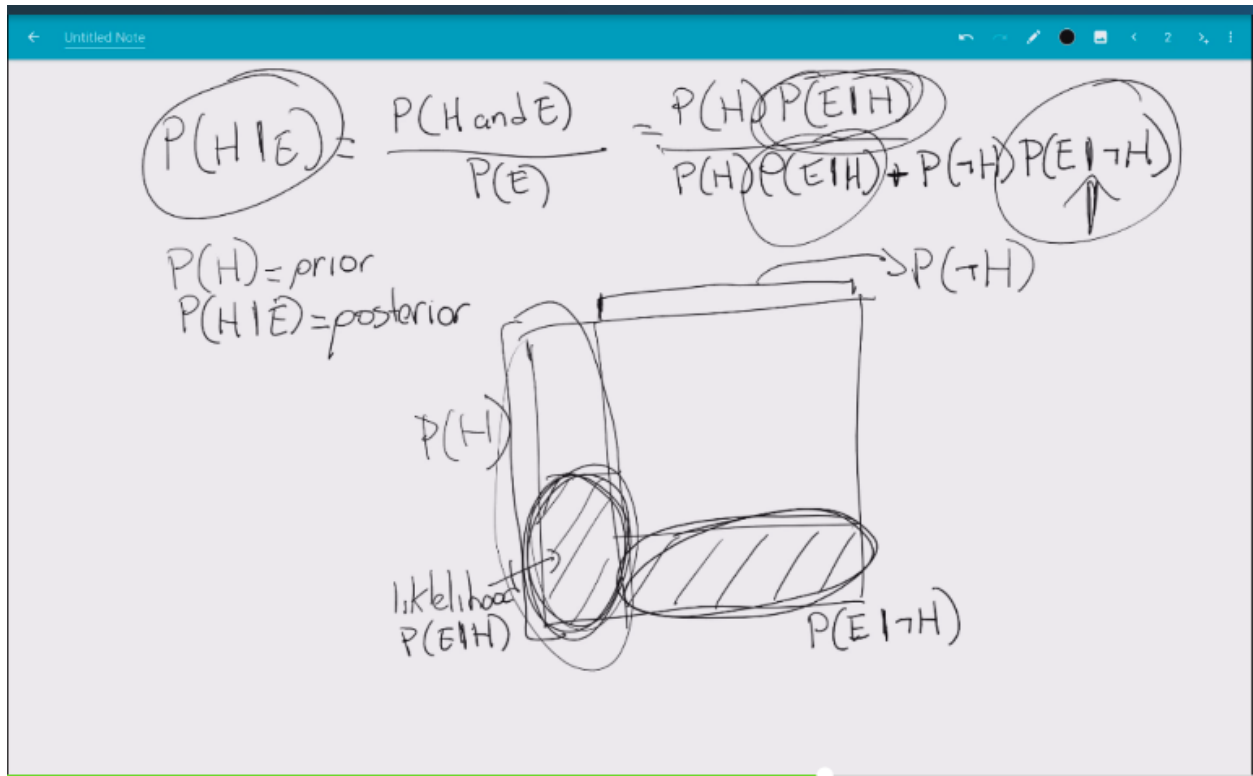
Teorema de Bayes

Thomas Bayes hijos de un sacerdote en Inglaterra en el siglo XVII.

Richar Price se topa con las notas de Bayes, que incorpora evidencia para actualizar nuestras creencias sobre algo que ocurra.

$P(\text{Hipotesis}) \Rightarrow$ se conoce como el prior, hipotesis antes de que recolectemos la evidencia

$P(H|E) \Rightarrow$ posterior teniendo ya evidencia para crear una cierta creencia.



Garbage on, garbage out:

Primer error de pensamientos.

GIGO \Rightarrow que si le avientas basura va a escupir basura. Conclusiones con respecto a datos que son incorrectos.

Ejemplos

Censo de 1840 en estados unidos, con métodos manuales los censos están plagados de errores. En este caso era un conjunto de datos basura.

Los esclavos se vuelven locos si los dejan ser libres. \Rightarrow error de hipótesis por datos basuras.

Cum Hoc Ergo Propter Hoc

Las correlación entre dos variables no significa causalidad.

Correlación positiva \Rightarrow las dos variables van hacia el mismo lugar

Correlación negativa \Rightarrow las dos variables van hacia diferente lugar

La correlación \neq no significa causalidad

Después de esto eso entonces a consecuencia de esto eso. \Rightarrow la forma de salir de esto es pensar que otras causas pueden estar detrás de este evento.

Prejuicio en el muestreo

Elimina la representatividad \Rightarrow generalizar el resultado en caso de que no sea representativa tu muestra.

Falacia del francotirador de Texas:

Una falacia que no toma en cuenta la aleatoriedad.


También sucede cuando solo nos enfocamos en las similitudes e ignoramos las diferencia.

Porcentajes Confusos

Una de las formas mas fáciles de mentir es con porcentajes.

Siempre se necesita contexto. \Rightarrow no se puede ver desde la abstraccion.

Siempre tenemos que ver el contexto y todos los datos para saber lo que se significan los porcentajes.

	Rendimiento 2018	Rendimiento 2019	Incremento	Incremento porcentaje
Escuela A	20	25 	5	25%
Escuela B	50	55	5	10%
Escuela C	95	100	5	5%

Desconfiar de los porcentajes en vacío, necesitan contexto.

Falacia de regresión:

Esta falacia consiste en que después de que sucede un evento extremo como podría el caso del casino de montecarlo, la gente toma acciones sin tener en cuenta que el evento tiende a regresar a la media. No es que el próximo evento sea menor e igual de extremo para equilibrar.

Con esto tenemos en claro que no podemos generar conclusiones en base a una regresión que puede fluctuar entre los datos.

Ejemplo ⇒ Un alumno que esta teniendo una mala racha se le aplica un castigo y entonces se da una regresión a la media por lo que el alumno saca conclusiones de que gracias al castigo mejoro sus notas.

Introduccion Machine Learning:

Es el campo de estudio que le da a las computadoras la habilidad de aprender sin ser explicitamente programadas.

Arthur Samuel 1959

Machine Learning se utiliza cuando:

programar un algoritmo de dato imposible

el problema es muy complejo o no se conocen algoritmos para resolverlo.

Ayuda a los humanos a encontrar patrones que no entendemos. (Data minig)

Aprendizaje supervisado vs Aprendizaje semi supervisado vs aprendizaje no supervisado

Batch vs Online Learning

Feature Vectors

Forma de agarrar rasgos de un objeto para ponerlo de forma numérica.

Lo que nosotros vemos en números.

Determinar los datos que importan o los que no importan.

Métricas de distancia

Forma de cuantificar que tan cercanos o lejanos están los vectores que estamos implementando.

Optimizar la distancia entre los vectores.

Caminos posibles en los que podemos implementar y llegar al destino.

Métricas de distancia:

- Distancia euclidiana: la raíz de el cuadrado de ambas distancias
- Distancia Manhattan: La diferencia de los valores absolutos y los suma

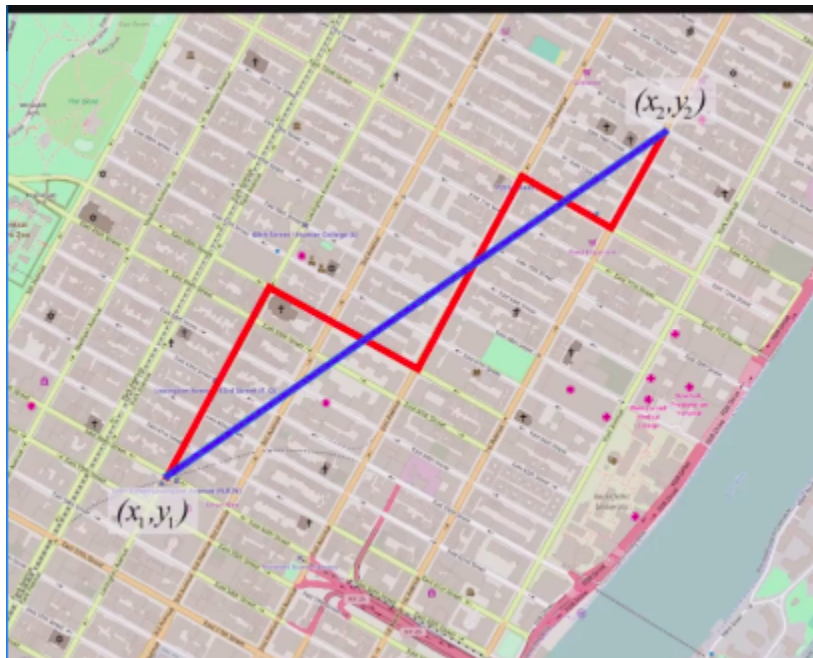
$$x = (a, b) \quad y = (c, d)$$

● Distancia euclidiana

$$\sqrt{(a - c)^2 + (b - d)^2}$$

● Distancia de Manhattan

$$|a - c| + |b - d|$$



<https://dataaspirant.com/five-most-popular-similarity-measures-implementation-in-python/>

Más métricas de distancia implementadas en python

Algoritmos de agrupamiento (clustering)

Clustering ⇒ permiten la estructura interna de los datos. Motores de búsqueda. Clasifica los vectores. Detección de riesgos.

Agrupar los datos. ⇒ estructurarlos cuando no sabes sus estructura y que la computadora nos muestre los resultados.

Agrupamiento Jerárquico

Tomar los puntos mas cercanos, agrupa los datos en un cluster y comparar la distancia con el cluster mas cercano agruparlo y repetirlo de manera iterativo.

Relaciones que existen en nuestro grupo.

Produce un dendograma ⇒ indica las relaciones que existen.

Agrupamiento K-means

funciona al asignar puntos al azar

k ⇒ determina el numero de grupos

medimos las distancias hacia todos los feature vectors

generamos la media en todo momento.

Centroide ⇒ la media del grupo

Repetimos el proceso hasta que ya no existan mejoras. Algoritmo convergió.

Escoger un numero de grupos correctos. ⇒ k

Para llegar a un agrupamiento que sea relevante.

Normalmente se usan muestras relevantes y muestras relativas.

ejemplo de manera gráfica ⇒ https://www.jacobsoft.com.mx/es_mx/k-means-clustering-con-python/

Agrupamiento estricto \Rightarrow cada uno de los puntos pertenecen a un grupo no hay puntos medios.

Agrupamiento Laxo (soft clustering) \Rightarrow en el cual en lugar de asignar un dato a un grupo se asigna probabilidades de pertenecer a alguno.

Técnicas de agrupamiento

Modelos conectivos \Rightarrow asumen que los puntos mas similares son aquellos que se encuentran mas cercanos dentro del espacio. Desventaja clara de estos tipos de modelos es que no escalan pero se puede usar una muestra representativa y aleatoria.

Modelos de centroide \Rightarrow estos modelos definen puntos medios a los cuales se determina la similitud en base a la cercanía con el centroide.

K-means es un ejemplo de este modelo.

Modelos de distribución \Rightarrow trata de asignar probabilidades de que cada dato pertenezca a una distribución específica o no. Por ejemplo normal binomial Poisson.

Modelos de densidad \Rightarrow Estos modelos analizan la densidad de los diferentes datos dividiendo en regiones o conjuntos. Luego se asigna de acuerdo con las áreas de densidad.

Introducción a la clasificación.

Determinar a que grupo pertenece un dato que no conocemos, ya tenemos información que nos ayuda a clasificar.

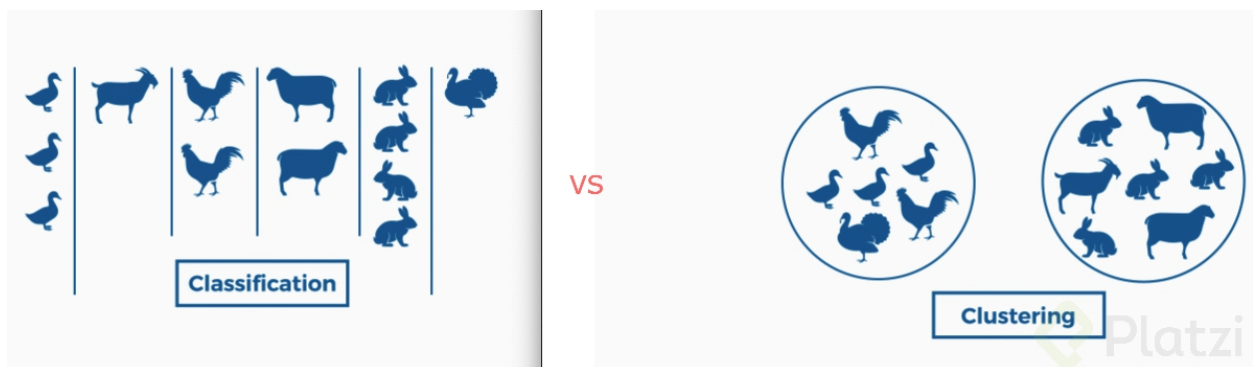
Un tipo de machine learning supervisado. Significa que para entrenar un modelo necesitamos un grupo de labels.

Generamos un modelo \Rightarrow clasificación de un dato que no tenemos etiquetas

aprendizaje semi supervisado \Rightarrow hay ciertos datos y preguntas que tienes que introducir.

aprendizaje no supervisado \Rightarrow genera clusters no clasificado

aprendizaje semi supervisado \Rightarrow aprender a clasificar, generar un modelo y generar la clasificación.



Key nearest neighbors

$k \Rightarrow$ significan los vecinos que vamos a utilizar para clasificar

Determinar a que numero pertenece nuestro data point.

$K \Rightarrow$ que sea par

Computacionalmente muy pesado y costoso

Clasificadores Lineales

Dividen un conjunto de datos con una linea (puede ser multidimensional), con esto generamos áreas dentro de nuestro espacio de búsqueda para que clasifique el nuevo dato fácilmente.