Índice

1. Marco teórico

1.1. Álgebra relacional

Los cimientos del modelo relacional son el *álgebra relacional*, las operaciones del álgebra producen nuevas relaciones, que pueden manipularse también por medio de operaciones del álgebra mismo. Una secuencia de operaciones de álgebra relacional forma una *expresión de álgebra relacional* cuyo resultado es una relación que representa el resultado de una consulta (o solicitud de consulta) de base de datos. Estas operaciones pueden clasificar en dos grupos, operaciones de conjuntos* de la teoría de conjuntos matemáticos refiere a la operaciones *UNION*, *INTERSECTION*, *SET DIFFERENCE* (*MINUS*) y *CARTESIAN PRODUCT* (*CROSS PRODUCT*). El otro grupo consiste en operaciones específicas para bases de datos relacionales: *JOIN*, *SELECT* y *PROJECT*. Estas dos últimas, por operar con una sola relación, son también conocidas como *operaciones unarias*.

↑ (así se llaman? checar libro en l'espagnol) "set operations"

SELECT elige un subconjunto de tuplas de una relación que satisfacen la condición de selección

$$\sigma_{<\text{condición de selección}>}(\mathrm{R}) \tag{1}$$

donde σ denota el operador SELECT, y la condición de selección es una expresión booleana especificada en los atributos de la relación R.

PROJECT produce una nueva relación de atributos y tuplas duplicadas

$$\pi_{<\text{lista de atributos}>}(\mathrm{R}) \tag{2}$$

donde π es el denota el operador PROJECT y la lista de atributos es la sublista de atributos deseados de la relación R.

Las operaciones con dos relaciones reciben el nombre de *operaciones binarias*. Si las relaciones $R(A_1, A_2, \ldots, A_n)$ y $S(B_1, B_2, \ldots, B_n)$ son compatibles de unión (tienen el mismo grado n y $dom(A_i) = dom(B_i)$ para $1 \leq i \leq n$) podemos usarlas para definir las siguientes operaciones binarias:

- UNION: El resultado de esta operación se denota $R \cup S$, es una relación que incluye todas las tuplas que están en R, S o ambos, se eliminan duplicados.
- INTERSECTION: El resultado de esta operación se denota $R \cap S$, es una relación que incluye todas las tuplas que están en ambos R y S.
- SET DIFFERENCE: El resultado de esta operación se denota $R - S$, es una relación que incluye todas las tuplas que están en ambos R pero no en S.

El operador CARTESIAN PRODUCT, denotado $R \times S$ es la operación binaria que no requiere compatibilidad de unión y produce un nuevo elemento al combinar cada miembro (tupla) de cada relación conjunto) con cada otro miembro de la otra relación. El resultado de $R(A_1, A_2, \ldots, A_n)$ y $S(B_1, B_2, \ldots, B_m)$ es una relación Q con atributos $Q(A_1, A_2, \ldots, A_n, B_1, B_2, \ldots, B_m)$ (en ese orden) de grado $n + m$. El resultado Q tiene una tupla por cada combinación de tuplas de R y S.

JOIN $\bowtie$, es el operador utilizado para combinar tuplas relacionadas de dos relaciones en una sola tupla. Pertinente para procesar relaciones entre relaciones. Si tenemos dos relaciones $R(A_1, A_2, \ldots, A_n)$ y $S(B_1, B_2, \ldots, B_m)$ podemos escribir la operación JOIN como

$$R \bowtie_{<\text{condiciones de unión}>} S \tag{3}$$

el resultado de la unión es la relación Q con atributos $Q(A_1, A_2, \ldots, A_n, B_1, B_2, \ldots, B_m)$ (en ese orden). El resultado Q tiene una tupla por cada combinación de tuplas de R y S que satisfacen las condiciones de unión.

1.2. Modelo de base de datos relacional

El modelo relacional de base de datos, según Date [Dat12] consiste en cinco componentes:

1. Una colección de tipos escalares, pueden ser definidos por el sistema (INTEGER, CHAR, BOOLEAN, etc.) o por el usuario.

2. Un generador de tipos de relaciones y un intérprete para las relaciones mismas.

3. Estructuras para definir variables relacionales de los tipos generados.

4. Un operador para asignar valores de relación a dichas variables.

5. Una colección relacionalmente completa para obtener valores relacionales de otros valores relacionales mediante operadores relacionales genéricos (Álgebra relacional 1.1).

Es importante comenzar definiendo los tipos, ya que las relaciones se definen sobre ellos, según Date, los tipos son "en esencia un conjunto finito de valores nombrado-todos los valores posibles de alguna categoría específica, por ejemplo, todos los números enteros posibles, todos los caracteres string posibles, todos los teléfonos de proveedores posibles, todos los documentos XML posibles, todas las relaciones con cierta cabecera posibles(y así sucesivamente)" [Dat12].

Cada atributo de cada relación es definido como de un tipo. Los atributos son pares ordenados de combinaciones atributo-nombre/tipo-nombre y una tupla es un par ordenado de atributos. El modelo relacional también soporta varios tipos de llaves, que poseen las propiedades de unicidad, ninguna contiene dos tuplas distintas con el mismo valor e irreductibilidad, ningún subconjunto suyo es tiene unicidad. La llave foránea (*FK*) es una combinación o set se atributos FK en una relación $r2$ tal que se requiere que cada valor FK sea igual a algún valor de alguna llave K en alguna relación $r1$ ($r1$ y $r2$ no son necesariamente distintos).

Una restricción de integridad (*constraint*) es una expresión booleana que debe evaluarse como verdadera. Los constraints de tipo definen los valores que constituyen un tipo dado, mientras que los constraints de base de datos limitan los valores que pueden aparecer en cierta base de datos. Las bases de datos suelen tener múltiples constraints específicos, expresados en términos de sus relaciones, sin embargo, el modelo relacional incluye dos constraints genéricos, que aplican a cada base de datos:

- Regla de integridad de identidad: Las llaves primarias no pueden ser nulas (*null*).
- Regla de integridad de referencia: No debe haber valores FK sin relación (si ***B*** referencia a ***A***, ***A*** debe existir).

Atributos {

código	(CHAR)	fecha	(DATE)	estado	(INT)
MX01		31-07-99		0	
MX02		30-07-99		0	
MX03		31-07-99		0	

Atributos {

código	CHAR	fecha	DATE	estado	INT
MX01		31-07-99		0	
MX02		30-07-99		0	
MX03		31-07-99		0	

1.3. Inteligencia artificial

Inteligencia artificial es definida como "el esfuerzo por automatizar tareas intelectuales normalmente realizadas por humanos"[Cho18], de este campo general se desprenden el *aprendizaje maquinal* y el *aprendizaje profundo*.

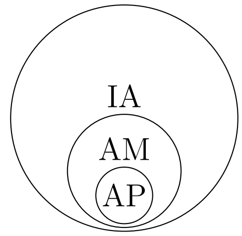


Figura 1: Aprendizaje profundo (AP), es un subcampo del aprendizaje maquinal (AM), que a su vez es un subcampo de la inteligencia artificial (IA)[Cho18].

1.4. Aprendizaje maquinal

La definición de aprendizaje maquinal de Tom Mitchell[Mit97] dice que “un programa de computadora aprende de experiencia $\boldsymbol{E}$ con respecto a una tarea $\boldsymbol{T}$ y una medición de rendimiento $\boldsymbol{P}$, si su rendimiento en $\boldsymbol{T}$, medido por $\boldsymbol{P}$, mejora con experiencia $\boldsymbol{E}$.”

Esto dignifica que a diferencia del paradigma clásico de programación, donde los humanos introducen órdenes y datos para ser procesados de acuerdo con dichas reglas, en el aprendizaje maquinal el humano introduce datos y respuestas esperadas de estos datos “y el producto son las reglas”*..

← “and out come the rules”

Si no es programado explícitamente, entonces un sistema de aprendizaje maquinal es entrenado: se le presentan muchos ejemplos relevantes a una tarea, y si encuentra una estructura estadística en ellos, genera reglas para automatizar la tarea.

1.5. Procesamiento de lenguaje natural

El *procesamiento de lenguaje natural*, es el conjunto de métodos para hacer accesible el lenguaje humano a las computadoras[Eis19].* Existen dos enfoques en lo que debe ser su tarea central:

← como que falta algo aquí?

- Entrenar sistemas de extremo a extremo* que transmuten texto sin procesar a cualquier estructura deseada.

← end-to-end, de principio a fin?

- Transformar texto en una pila de estructuras lingüísticas de uso general que en teoría deben poder soportar cualquier aplicación.

Dos de los módulos básicos de NLP son *búsqueda* y *aprendizaje* con los que se puede resolver muchos problemas que podemos describir en la siguiente forma matemática

$$\hat{y} = \underset{y \in Y(x)}{argmax} \Psi(x, y; 0), \tag{4}$$

donde,

- x es la entrada, un elemento de un conjunto X.
- y es el resultado, un elemento de un conjunto Y.
- Ψ es una función de puntuación (también conocida como *modelo*), que va desde el conjunto $X \times Y$ hasta los números reales.
- $\emptyset$ es el vector de parámetros para $\boldsymbol{\Psi}$.
- $\hat{y}$ es el resultado previsto, que es elegido para maximizar la función de puntuación.

El módulo de búsqueda se encarga de computar el $argmax$ de la función $\boldsymbol{\Psi}$, es decir, encuentra el resultado $\hat{y}$ con la mejor puntuación con respecto a la entrada x. El módulo de aprendizaje encuentra los parámetros θ por medio del procesamiento de grandes conjuntos de datos de ejemplos etiquetados $\{(x^i, y^i)\}_{i=1}^{N}$.

1.6. Regresión

Es común en el procesamiento de lenguaje natural utilizar *modelos estadísticos lineales* para el modelado* del valor medio de una variable de resultado Y como función de una independiente que pueden definirse con el modelo de *regresión lineal simple*

↑ modelo modelado?

$$E(Y) = \beta_0 + \beta_1 x \tag{5}$$

en cambio, cuando se tiene más de una variable independiente se utiliza el modelo de *regresión lineal múltiple*

$$E(Y) = \beta_0 + \beta_1 x + \ldots + \beta_k x_k \tag{6}$$

nada de esto sirve

1.7. Covarianza y correlación

La covarianza mide qué tanto o tan poco las dos variables aleatorias cuyos valores esperados existen y son positivos tienen dependencia lineal, denotada $cov(X, Y)$ la covarianza de X y Y es definida como

$$cov(X,Y) = E[(X - EX)(Y - EY)] \tag{7}$$

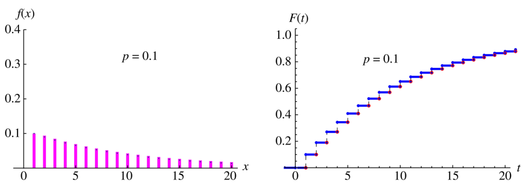
Cuando no se tiene una referencia para usar la covarianza, tiene sentido escalarla de acuerdo a la desviación estándar de las variables[Mat17]

$$p(X,Y) = \frac{cov(X,Y)}{\sqrt{var(X)}\sqrt{var(Y)}} \tag{8}$$

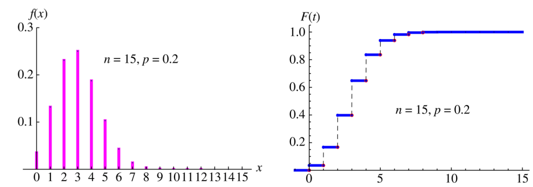
denotada $corr(X,Y)$ de esta es la *correlación* de X y Y. Un coeficiente de relación $p = 0$ indica que no hay relación.

1.8. Distribuciones

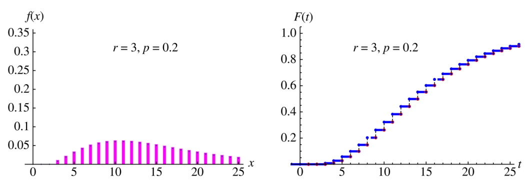
Blablabla distribuciones comunes para variables discretas:[BKP20]
SORT ME AND EXPLAIN



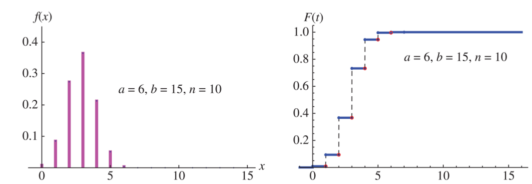
(a) Geométrica



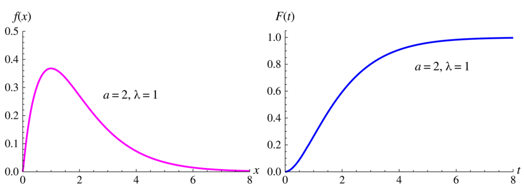
(b) Binominal



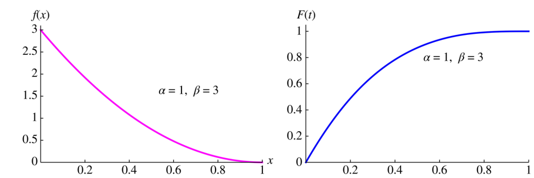
(c) Negativa binominal



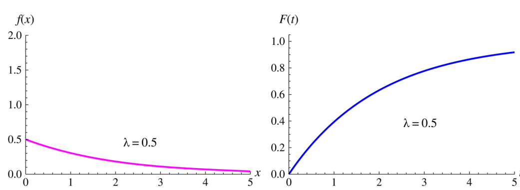
(d) Hipergeométrica



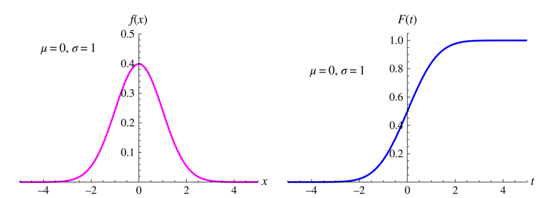
(e) Γ



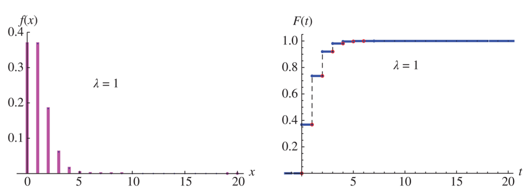
(f) Beta



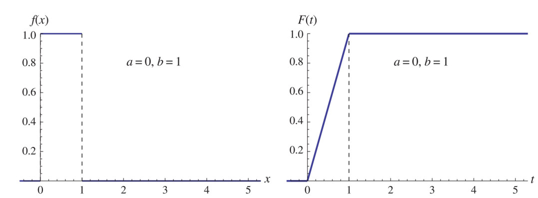
(g) $\epsilon(\lambda)$



(h) Normal



(i) Poisson



(j) Uniforme

1.9. Probabilidad condicional

Mientras que las distribuciones anteriores nos han dado toda la información acerca de la probabilidad de las variables aleatorias, cuando sólo se requiere un número que extraiga su valor, podemos utilizar la *media*, también conocida como *valor esperado*.

Dada una lista de números $x_1, x_2 \ldots, x_n$, para obtener la *media aritmética*, estos se suman y dividen entre n:

$$\bar{x} = \frac{1}{n} \sum_{j=1}^{n} x_j, \tag{9}$$

la *media ponderada* de $x_1, x_2 \ldots, x_n$ se obtiene de la siguiente forma:

$$\text{media ponderada}(x) = \frac{1}{n} \sum_{j=1}^{n} x_j P_j, \tag{10}$$

donde los pesos $p_1, p_2 \ldots, p_n$ son números no negativos previamente especificados que suman a 1.

El valor esperado o media de una variable aleatoria discreta X cuyos posibles valores distintos son $x_1, x_2 \ldots$ es definida por

$$E(X) = \sum_{j=1}^{\infty} x_j P(X = xj), \tag{11}$$

si el soporte es finito, entonces se reemplaza por una suma finita, escribiéndose de la siguiente forma:

$$E(X) = \sum_{x} \underbrace{x}_{\text{valor}} \underbrace{P(X = x)}_{\substack{\text{Función de} \\ \text{probabilidad} \\ \text{en } x}} . \tag{12}$$

***********nuevo * Varianza

$$Var(X) = E(X - EX)^2 = E(X^2) - (EX)^2 \tag{13}$$

***********nuevo

El valor esperado de una suma de variables aleatorias es la suma de sus valores esperados individuales, este es el teorema de la *linealidad del valor esperado*, donde para cada variable aleatoria X, Y y cada constante c,

$$\begin{gathered} E(X + Y) = E(X) + E(Y), \\ E(cX) = cE(X). \end{gathered} \tag{14}$$

1.9.1. Binominal geométrica y negativa

Distribución geométrica: Se tiene una secuencia de ensayos independientes Bernoulli, cada uno con la misma probabilidad de éxito $p \in (0,1)$, con ensayos realizados hasta que se alcanza el éxito. X es el número de *fallas* antes de la primera prueba exitosa por lo que X tiene una *distribución geométrica* con un parámetro p; denotado $X \sim Geom(p)$. Con esto podemos llegar a los teoremas de *distribución geométrica de la función de probabilidad*, cuando $X \sim Geom(p)$, entonces la función de probabilidad de X será

$$P(X = k) = q^k p \tag{15}$$

para $k = 1, 2, \ldots$, cuando $q = 1 - p$; y el teoremas de *distribución geométrica de la función de distribución acumulativa*, cuando $X \sim Geom(p)$, entonces la función de distribución acumulativa de X será

$$F(x) = \begin{cases} 1 - q^{\lfloor x \rfloor + 1}, \text{ si } x \geq 0; \\ 0, \text{ si } x < 0, \end{cases} \tag{16}$$

cuando $q = 1 - q$ y $\lfloor x \rfloor$ es el mayor entero y menor o igual a x.

El valor esperado geométrico de $X \sim Geom(p)$ es

$$E(X) = \sum_{k=0}^{\infty} k q^k p, \tag{17}$$

cuando $q = 1 - p$. Aunque esta no es una serie geométrica, podemos llegar a ello

$$\begin{gathered} \textstyle\sum_{k=0}^{\infty} q^k = \frac{1}{1-q} \\ \textstyle\sum_{k=0}^{\infty} k q^{k-1} = \frac{1}{1-q^2}, \end{gathered} \tag{18}$$

finalmente multiplicamos ambos lados por pq, recuperando la suma original que queríamos encontrar

$$E(X) = \sum_{k=0}^{\infty} k q^k p = pq \sum_{k=0}^{\infty} k q^{k-1} = pq \frac{1}{(1-q)^2} = \frac{q}{p}. \tag{19}$$

Primer valor esperado de éxito *FS*, podemos definir a $Y \sim FS(p)$ como $Y = X + 1$ donde $X \sim Geom(p)$, por lo que tenemos

$$E(Y) = E(X+1) = \frac{q}{p} + 1 = \frac{1}{p}. \tag{20}$$

Las *distribuciones binominales negativas* generalizan la distribución geométrica en lugar de esperar por un éxito, podemos esperar por cualquier número predeterminado r de éxitos. En una secuencia de ensayos independientes Bernoulli con probabilidad de éxito p, si X es el número de *fallas* antes del éxito número r, entonces se dice que Xtiene una distribución binominal negativa con parámetros r y p, denotado $X \sim NBin(r, p)$.

La distribución binominal cuenta el número de éxitos en un número fijo de ensayos, mientras que la binominal negativa cuenta el número de fallas hasta alcanzar cierto número de éxitos. Si $X \sim NBin(r, p)$, entonces la función de probabilidad de X es

$$P(X = n) = \binom{n + r - 1}{r - 1} p^r q^n \tag{21}$$

para $n = 0, 1, 2 \ldots$, donde $q = 1 = p$.

1.9.2. Varianza, agregarlas a la secc que pertenecen

Varianza de la geométrica (agregarla a la geom después de editar todo y hacerlo más breve)

$$Var(X) = E(X^2) - (EX)^2 = \frac{q(1+q)}{p^2} - (\frac{q}{p})^2 = \frac{q}{p^2} \tag{22}$$

Varianza de la geométrica (lo mismo)

$$Var(X) = E(X^2) - (EX)^2 = (n(n-1)p^2 + np) - (np)^2 = np(1-p). \tag{23}$$

Una variable aleatoria X tiene *distribución de Poisson* (denotada $X \sim Pois(\lambda)$) con el parámetro λ, cuando $\lambda > 0$ si la PMF de x es

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, k = 0, 1, 2, \ldots. \tag{24}$$

Varianza de la distribución de Poisson es

$$Var(X) = E(X^2) - (EX)^2 = \lambda(1+\lambda) - \lambda^2 = \lambda \tag{25}$$

1.9.3. r.v. continuas

A diferencia de las variables discretas, las *variables aleatorias continuas* pueden tomar cualquier valor real en un intervalo y tienen una *distribución continua.* Para obtener la probabilidad deseadaWHOMST, se debe integrar la función de densidad de probabilidad sobre el rango apropiado

$$P(X \in A) = \int_A f(x)dx \tag{26}$$

La distribución logística se obtiene

$$F(x) = \frac{e^x}{1+e^x}, x \in \Re \tag{27}$$

El valor esperado de la continua función de distribución acumulada f es

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx \tag{28}$$

la continua U tiene dist unif en el intervalo (a,b), denotada $U \sim Unif(a,b)$ si el área acumulada bajo la función de densidad de probabilidad es

$$f(x) = \begin{cases} \frac{1}{b-a}, \text{ si } a < x < b; \\ \text{de lo contrario, 0} \end{cases} \tag{29}$$

$U \sim Unif(a,b)$ de una variable aleatoria es

$$\tilde{U} = a + (b-a)U \tag{30}$$

Su varianza es

$$Var(U) = \frac{(b-a)^2}{12} \tag{31}$$

La distribución normal ($X \sim N(\mu, \sigma^2)$)cuando $\mathbb{Z} \sim N(0,1)$ es

$$X = \mu + \sigma\mathbb{Z} \tag{32}$$

Por lo tanto obtendremos el valor esperado de μ y varianza de σ^2

$$X = (\mu + \sigma\mathbb{Z}) = E(\mu) + \sigma E(\mathbb{Z}) = \mu \tag{33}$$

$$Var = (\mu + \sigma\mathbb{Z}) = Var(\sigma\mathbb{Z}) = \sigma^2 Var(\mathbb{Z}) = \sigma^2. \tag{34}$$

La *distribución exponencial* de X con un parámetro λ, cuando $\lambda > 0$ si su función de densidad de probabilidad es $f(x) = \lambda e^{-\lambda x}, x > 0$; denotada como $X \sim Expo(\lambda)$ es la siguiente

$$F(x) = 1 - e^{-\lambda x}, x > 0 \tag{35}$$

1.10. Variables aleatorias y sus distribuciones

Una *variable aleatoria* es una función asignando un número real $\mathbb{R}$ a cada posible resultado de un experimento. Con una muestra en espacio S, una variable aleatoria X asigna el valor numérico $X(s)$ a cada resultado posible s del experimento. La aleatoriedad viene del hecho que tenemos un experimento aleatorio (con probabilidades descritas por la función de probabilidad P). Las variables aleatorias simplifican la notación y expanden la habilidad de cuantificar y resumir resultados de experimentos.

Se dice que una variable X es discreta cuando si hay una lista finita de valores $a_,a_2, \dots, a_n$ o un una lista infinita de valores $a_,a_2, \dots$ de tal forma que $P(X = a_j$ para algún $j) = 1$. Si X es una variable aleatoria discreta, entonces el conjunto infinito o contable de valores x tal que $P(X = x)$ se llama *soporte* de X. En contraste una variable aleatoria continua puede tomar cualquier valor real en un intervalo.

1.10.1. Función de probabilidad

La forma más natural de expresar la distribución de variables aleatorias discretas es la *función de probabilidad*[BH19]* que, para una X discreta, es la función p_X dada por $p_X(x) = P(X = x)$. El teorema de *funciones de probabilidad válidas* dice que cuando X es una variable aleatoria con soporte $x1, x2, \dots$, la función de probabilidad p_X de x debe satisfacer los siguiente criterios:

↑ PMF, por sus siglas en inglés?

- No negativo $p_X(x) > 0$ si $x = x_j$ para un j, y $p_X(x) = 0$, de otra forma;
- Suma 1: $\sum_{j=1}^{\infty} p_X(x_j) = 1$.

el primer criterio es verdadero porque la probabilidad es no negativa, el segundo es verdadero ya que X debe tomar *algún* valor, y los eventos $X = xj$ están disjuntos, entonces

$$\sum_{j=1}^{\infty} P(X = x_j) = P\left(\bigcup_{j=1}^{\infty} \{X = x_j\}\right) = P(X = x_1 \text{ ó } X = x_2 \text{ ó } \dots) = 1. \tag{36}$$

1.10.2. Distribución de Bernoulli y binominal

Una variable aleatoria tiene la *distribución de Bernoulli* con un parámetro p si $P(X = 1) = p$ y $P(X = 0 = 1 - p)$, cuando $0 < p < 1$. Se escribe como $X \sim Bern(p)$, el símbolo $\sim$ significa "distribuido como" y la probabilidad p es el *parámetro*, que determina qué distribución de Bernoulli específica tenemos.

Supóngase que se realizan n ensayos Bernoulli independientes, cada uno con probabilidad p de éxito. X sea el número de éxitos, la distribución X se llama *distribución* *binominal* con parámetros n y p; se escribe $X \sim Bin(p, n)$. $Bern(p)$ es la misma distribución que $Bin(1, p)$. Bernoulli es un caso especial de binominal, si $x \sim Bin(1, p)$, entonces la función de probabilidad de X es

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \tag{37}$$

para $k = 0, 1, \ldots, n$ (y por otra parte $P(X = k) = 0$).

1.10.3. Distribución de hipergeométrica

Si $X \sim HGeom(w, b, n)$, entonces la función de probabilidad de X es

$$P(X = k) = \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}, \tag{38}$$

para enteros k satisfaciendo $0 \leq k \leq w$ y $0 \leq n-k \leq b$, y $P(X = k) = 0$. La estructura esencial de la distribución hipergeométrica se basa en que objetos en su población están clasificados usando dos tipos de etiquetas, al menos una de estas siendo asignada al azar. Las distribuciones $HGeom(w, b, n)$ y $HGeom(n, w+b-n, 1)$ son idénticas si X y Y tienen la misma distribución, podemos demostrarlo algebraicamente:

$$P(X = k) = \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}} = \frac{w!b!n!(w + b - n)!}{k!(w + b)!(w - k)!(n - k)!(b - n + k)!} \tag{39}$$

$$P(X = k) = \frac{\binom{n}{k}\binom{w+b-n}{w-k}}{\binom{w+b}{w}} = \frac{w!b!n!(w + b - n)!}{k!(w + b)!(w - k)!(n - k)!(b - n + k)!}. \tag{40}$$

1.10.4. Distribución uniforme discreta

Teniendo C, un conjunto finito no vacío de números, se elige un número uniformemente al azar (o sea que todos los números tienen la misma posibilidad de ser elegidos), llámese X. Entonces se dice que X una *distribución uniforme discreta* con el parámetro C. Se dice entonces que la función de probabilidad de $X \sim DUNif(C)$ (la distribución uniforme discreta de X) es

$$P(X = x) = \frac{1}{|C|} \tag{41}$$

para $x \in C$ (de lo contrario 0) ya que la función de probabilidad debe sumar 1.

1.10.5. Función de distribución acumulada

Esta función describe la distribución de todas las variables aleatorias (a diferencia de la función de probabilidad que sólo se aplica a las discretas). La *función de distribución acumulada* de una variable aleatoria X es la función F_X dada por $F_X(x) = P(X \leq x)$ y tiene las siguientes propiedades:

- Incrementos: Si $x_1 \leq x_2$, then $F(x_1) \leq F(x_2)$.
- Continua por la derecha: Es continua por la posibilidad de tener saltos. Cuando hay saltos es continua por la derecha, es decir, por cada a se tiene

$$F(a) = \lim_{c \to a^+} F(x). \tag{42}$$

- Convergencia de 0 y 1 en los límites

$$\lim_{x \to \infty} F(x) = 0 \ \text{ y } \lim_{x \to \infty} F(x) = 1. \tag{43}$$

1.11. Valor esperado

Mientras que las distribuciones anteriores nos han dado toda la información acerca de la probabilidad de las variables aleatorias, cuando sólo se requiere un número que extraiga su valor, podemos utilizar la *media*, también conocida como *valor esperado*. Dada una lista de números $x_1, x_2 \ldots, x_n$, para obtener la *media aritmética*, estos se suman y dividen entre n:

$$\bar{x} = \frac{1}{n} \sum_{j=1}^{n} x_j, \tag{44}$$

la *media ponderada* de $x_1, x_2 \ldots, x_n$ se obtiene de la siguiente forma:

$$\text{media ponderada}(x) = \frac{1}{n} \sum_{j=1}^{n} x_j P_j, \tag{45}$$

donde los pesos $p_1, p_2 \ldots, p_n$ son números no negativos previamente especificados que suman a 1.

El valor esperado o media de una variable aleatoria discreta X cuyos posibles valores distintos son $x_1, x_2 \ldots$ es definida por

$$E(X) = \sum_{j=1}^{\infty} x_j P(X = xj), \tag{46}$$

si el soporte es finito, entonces se reemplaza por una suma finita, escribiéndose de la siguiente forma:

$$E(X) = \sum_x \underbrace{x}_{\text{valor}} \underbrace{P(X = x)}_{\substack{\text{Función de} \\ \text{probabilidad} \\ \text{en } x}} . \tag{47}$$

***********nuevo * Varianza

$$Var(X) = E(X - EX)^2 = E(X^2) - (EX)^2 \tag{48}$$

***********nuevo

El valor esperado de una suma de variables aleatorias es la suma de sus valores esperados individuales, este es el teorema de la *linealidad del valor esperado*, donde para cada variable aleatoria X, Y y cada constante c,

$$\begin{gathered} E(X+Y) = E(X) + E(Y), \\ E(cX) = cE(X). \end{gathered} \tag{49}$$

1.11.1. Binominal geométrica y negativa

Distribución geométrica: Se tiene una secuencia de ensayos independientes Bernoulli, cada uno con la misma probabilidad de éxito $p \in (0, 1)$, con ensayos realizados hasta que se alcanza el éxito. X es el número de *fallas* antes de la primera prueba exitosa por lo que X tiene una *distribución geométrica* con un parámetro p; denotado $X \sim Geom(p)$. Con esto podemos llegar a los teoremas de *distribución geométrica de la función de probabilidad*, cuando $X \sim Geom(p)$, entonces la función de probabilidad de X será

$$P(X = k) = q^k p \tag{50}$$

para $k = 1, 2, \ldots$, cuando $q = 1 - p$; y el teoremas de *distribución geométrica de la función de distribución acumulativa*, cuando $X \sim Geom(p)$, entonces la función de distribución acumulativa de X será

$$F(x) = \begin{cases} 1 - q^{\lfloor x \rfloor + 1}, \text{ si } x \geq 0; \\ 0, \text{ si } x < 0, \end{cases} \tag{51}$$

cuando $q = 1 - q$ y $\lfloor x \rfloor$ es el mayor entero y menor o igual a x.

El valor esperado geométrico de $X \sim Geom(p)$ es

$$E(X) = \sum_{k=0}^{\infty} kq^k p, \tag{52}$$

cuando $q = 1 - p$. Aunque esta no es una serie geométrica, podemos llegar a ello

$$\begin{aligned} \textstyle\sum_{k=0}^{\infty} q^k &= \tfrac{1}{1-q} \\ \textstyle\sum_{k=0}^{\infty} kq^{k-1} &= \tfrac{1}{1-q^2}, \end{aligned} \tag{53}$$

finalmente multiplicamos ambos lados por pq, recuperando la suma original que queríamos encontrar

$$E(X) = \sum_{k=0}^{\infty} kq^k p = pq \sum_{k=0}^{\infty} kq^{k-1} = pq \frac{1}{(1-q)^2} = \frac{q}{p}. \tag{54}$$

Primer valor esperado de éxito *FS*, podemos definir a $Y \sim FS(p)$ como $Y = X + 1$ donde $X \sim Geom(p)$, por lo que tenemos

$$E(Y) = E(X+1) = \frac{q}{p} + 1 = \frac{1}{p}. \tag{55}$$

Las *distribuciones binominales negativas* generalizan la distribución geométrica en lugar de esperar por un éxito, podemos esperar por cualquier número predeterminado r de éxitos. En una secuencia de ensayos independientes Bernoulli con probabilidad de éxito p, si X es el número de *fallas* antes del éxito número r, entonces se dice que Xtiene una distribución binominal negativa con parámetros r y p, denotado $X \sim NBin(r, p)$.

La distribución binominal cuenta el número de éxitos en un número fijo de ensayos, mientras que la binominal negativa cuenta el número de fallas hasta alcanzar cierto número de éxitos. Si $X \sim NBin(r, p)$, entonces la función de probabilidad de X es

$$P(X = n) = \binom{n + r - 1}{r - 1} p^r q^n \tag{56}$$

para $n = 0, 1, 2\ldots,$ donde $q = 1 = p$.

1.11.2. Varianza, agregarlas a la secc que pertenecen

Varianza de la geométrica (agregarla a la geom después de editar todo y hacerlo más breve)

$$Var(X) = E(X^2) - (EX)^2 = \frac{q(1+q)}{p^2} - (\frac{q}{p})^2 = \frac{q}{p^2} \tag{57}$$

Varianza de la geométrica (lo mismo)

$$Var(X) = E(X^2) - (EX)^2 = (n(n-1)p^2 + np) - (np)^2 = np(1-p). \tag{58}$$

Una variable aleatoria X tiene *distribución de Poisson* (denotada $X \sim Pois(\lambda)$) con el parámetro λ, cuando $\lambda > 0$ si la PMF de x es

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, k = 0, 1, 2, \ldots. \tag{59}$$

Varianza de la distribución de Poisson es

$$Var(X) = E(X^2) - (EX)^2 = \lambda(1+\lambda) - \lambda^2 = \lambda \tag{60}$$

1.11.3. r.v. continuas

A diferencia de las variables discretas, las *variables aleatorias continuas* pueden tomar cualquier valor real en un intervalo y tienen una *distribución continua.* Para obtener la probabilidad deseadaWHOMST, se debe integrar la función de densidad de probabilidad sobre el rango apropiado

$$P(X \in A) = \int_A f(x)dx \tag{61}$$

La distribución logística se obtiene

$$F(x) = \frac{e^x}{1+e^x}, x \in \Re \tag{62}$$

El valor esperado de la continua función de distribución acumulada f es

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx \tag{63}$$

la continua U tiene dist unif en el intervalo (a, b), denotada $U \sim Unif(a, b)$ si el área acumulada bajo la función de densidad de probabilidad es

$$f(x) = \begin{cases} \frac{1}{b-a}, \text{ si } a < x < b; \\ \text{de lo contrario, } 0 \end{cases} \tag{64}$$

$U \sim Unif(a, b)$ de una variable aleatoria es

$$\tilde{U} = a + (b - a)U \tag{65}$$

Su varianza es

$$Var(U) = \frac{(b - a)^2}{12} \tag{66}$$

La distribución normal ($X \sim N(\mu, \sigma^2)$)cuando $\mathbb{Z} \sim N(0, 1)$ es

$$X = \mu + \sigma\mathbb{Z} \tag{67}$$

Por lo tanto obtendremos el valor esperado de μ y varianza de σ^2

$$X = (\mu + \sigma\mathbb{Z}) = E(\mu) + \sigma E(\mathbb{Z}) = \mu \tag{68}$$

$$Var = (\mu + \sigma\mathbb{Z}) = Var(\sigma\mathbb{Z}) = \sigma^2 Var(\mathbb{Z}) = \sigma^2. \tag{69}$$

La *distribución exponencial* de X con un parámetro λ, cuando $\lambda > 0$ si su función de densidad de probabilidad es $f(x) = \lambda e^{-\lambda x}, x > 0$; denotada como $X \sim Expo(\lambda)$ es la siguiente

$$F(x) = 1 - e^{-\lambda x}, x > 0 \tag{70}$$

2. Objetivos

Comenzar desde cero cada proyecto es ineficiente. (esto es temporal)

2.1. General

Diseñar y desarrollar una
plataforma/framework de ML para NLP
reutilizable y/o de uso general/ ¿para casos de uso similares?

2.2. Particulares

Revisé algunas tesis para darme la idea, necesito saber un poco más del tema para desarrollar esta parte, creo que sería algo así:

1. Investigación
2. Análisis info
3. Diseño/desarrollo
4. Comprobación

2.3. Hipótesis

Al integrar
/tecnologías/técnicas/modelos/librerías/frameworks/
de RDB, ML, NLP, ¿transfer learning?, ¿deep learning?,...
se de puede /diseñar/desarrollar/, ¿e implementar? una
/plataforma/framework/
/genérica/normalizada/universal/reutilizable/estandarizada/compatible
/con/en/para/ /casos de uso similares/diversos casos de uso/
(/reduciendo tiempo/ahorrando recursos/.) estas oración se sale del scope

3. Metas

Contribuir por medio del diseño y desarrollo de la
/plataforma/framework/ de ML para NLP
que será utilizable en variedad de casos reales /con características similares./

4. Metodologías

5. Referencias

[BH19] Joseph K. Blitzstein y Jessica Hwang. *Introduction to Probability [Introducción a la probabilidad]*. 2.ª ed. Texts in Statistical Science [Textos en ciencia estadistica]. Florida: CRC Press, 2019. ISBN: 9781138369917.

[BKP20] N. Balakrishnan, Markos V. Koutras y Konstantinos G. Politis. *Introduction to probability: models and applications [Introducción a la probabilidad: modelos y aplicaciones]*. Hoboken: John Wiley & Sons, Inc., 2020. ISBN: 9781118123348.

[Cho18] François Chollet. *Machine Learning With Python [Machine Learning]*. New York: Manning Publications Co., 2018. ISBN: 9781617294433.

[Dat12] C. J. Date. *SQL and Relational Theory: How to Write Accurate SQL Code [SQL y teoria relacional: Cómo escribir código SQL correcto]*. 2.ª ed. California: 0'Reilly Media, 2012. ISBN: 9781449316402.

[Eis19] Jacob Eisenstein. *Introduction to natural language processing [Introducción al procesamiento de lenguaje natural]*. Adaptive computation and machine learning [Computación adaptativa y aprendizaje maquinal]. Cambridge: MIT Press, 2019. ISBN: 9780262042840.

[Mat17] Norman S Matloff. *Statistical regression and classification from linear models to machine learning [Regresión estadística y clasificación de modelos lineales a aprendizaje maquinal]*. FLorida: CRC Press, 2017. ISBN: 9781138066565.

[Mit97] Tom Mitchell. *Machine learning [Aprendizaje maquinal]*. New York: MacGraw - Hill, 1997. ISBN: 0070428077.