Documentación

Javier Andrés Causil Martínez

Universidad de Antioquia - Científico de datos

Portafolio

Causil

Índice general

Índice de figuras

Eureka infinity es mi proyecto personal, su finalidad es publicar todo lo relacionado con la matemática que yo vaya produciendo relacionado con mi entorno, ya se han cursos, avances de proyectos personales, hasta hoy día tengo poca experiencia profesional, pero con el tiempo, la disciplina y la constancia en el estudio, estaré creciendo gracias a la comunidad.

Parte I

Ciencia de datos

¿Qué es data ciencia? Es el proceso de descubrir información valiosa de los datos.

¿Cuál es su finalidad?

1. Tomar decisiones y crear estrategias de negocio.
2. Crear productos de software más inteligentes y funcionales.

¿De que trata este proceso?:

1. Obtención de los datos: a travéz de encuestas
2. Transformar y limpiar los datos.
3. Explorar, analizar y visualizar datos.
4. Usar modelos de machine learning*.
5. Integrar datos e IA a productos de software.

¿Qué es Data Science?

Data science o ciencia de datos es el proceso de descubrir información valiosa de los datos.

¿Cuál es su finalidad? Tomar decisiones y crear estrategias de negocio Crear productos de software más inteligentes y funcionales.

¿De qué trata este proceso?

Obtención de los datos Mediciones Encuestas Internet

Transformar y limpiar los datos Incompletos Formato Incorrecto

Explorar, analizar y visualizar datos Patrones o tendencias Insights Visualizaciones, gráficos o reportes

Usar modelos de machine learning

Machine learning o aprendizaje automático es una rama de inteligencia artificial. Su objetivo es que las computadoras aprendan. En machine learning, las computadoras observan grandes cantidades de datos y construyen un modelo capaz de generar predicciones para resolver problemas.

Integrar datos e IA a productos de software Ponerlos a disposición del usuario final.

La ciencia de datos es una intercepción de conocimiento entre (matemáticas y estadística), (ciencias computacionales) y conocimiento del dominio.

Capítulo 1

Proyectos data Analysis

Poner en practica lo mas rápido que se pueda, tener proyectos personales, en que gasto los dineros del mes, que productos consigo cada mes, encontrar anomalías, proyectos con kaggle.

Capítulo 2

Análisis de datos

¿Qué es ciencia de datos y big data? ¿Cómo afectan a mi negocio?

"¿Qué haces en tu trabajo (como científico de datos)? Mi trabajo es crear una solución matemática o estadística para un problema del negocio"

2.1. ¿Qué tipo de información podemos analizar?

Descubrir qué tipos de información existen, qué industrias los usan y qué tipo de acciones podemos tomar a partir de ellos.

Los principales datos que existen son:

Personas: Este tipo de datos lo extraemos de las personas, es decir lo generamos nosotros cuando le damos like a una foto de facebook, de preferencia, tipo de videos, de quien te gusta mas el contenido, subiendo una foto y etiquetando a un compañero.

Transacciones: las monetarias y las no monetarias, cualquier transación que hago con una tarjeta de crédito o débito, cuando hacemos un pago electrónico o dígital queda una huella, queda un registro de quien lo hizo, por que monto lo hizo y en que establecimiento lo hizo, es muy interesante por que las bancas digitales pueden hacerte recomendaciones sobre el tipo de comercio que te podía interesar.

No finacieras: las compañias telefónicas identifican cual es tu patrón habitual, cuantas llamadas haces, a que personas llamas, cuanto duran tus llamadas, y a partir de esto te llaman para que no abandones el servicio.

Navegación web: Estas son las famosas cookies, ellas están advirtiendo de la información que van a recoger.

Machine 2 machine: Una conexion de una maquina a otra maquina, la usan las plataformas de transporte, google maps y para hacer la locación entre dispositivos.

Biométricos: Cada vez son mas habituales y únicas, huellas digitales, reconocimiento facial.

2.2. Flujo de trabajo en ciencia de datos: fases, roles y oportunidades laborales

Roles en datos:

Ingeniero de datos: crear bases de datos Hacer que la empresa, hace la conexion de los dispositivos y las bases de datos,

Analista business intelligence: A partir de la información que ha creado el ingeniero de datos va extraer la data, crear cuadros de control, crear dashboard, monitoreo, va automatizar estos procedimientos para que cualquier persona de la empresa pueda interpretarla, las herramientas mas utilizadas son SQL y Excel. No necesariamente sabe Python.

Data Scientist: Sabe hacer el rol del analista, sabe extraer la información y sabe predecir, con las herramientas de estadística, nos guía a donde vamos.

Data Translator: Nos ayuda a proyectar el equipo, nos ayuda a comunicar con los otros equipos del negocio.

2.3. Herramientas para cada etapa del análisis de datos

El primero es el rol del analista y del ingeniero estas son las personas que crean bases de datos y utilizan SQL, se sintetiza la información de la base de datos.

El científico de datos son herramientas predictivas, son R y Python, R es mas estadístico análisis descriptivo,

2.4. Python en ciencia de Datos

Por que numpy para el anÃ¡lisis de datos. Tenemos tres cosas a destacar

1. Un poderoso objeto array multidimensional.
2. Funciones matem

Crear un virtual environments ejecutamos la siguiente linea de comando

```
python3 -m venv my_env
source bin/activate
```

Capítulo 3

Proyectos

3.1. Credit Card Fraud Detection

Anonymized credit card transactions labeled as fraudulent or genuine

About Dataset

Context

The dataset contains transactions made by credit cards in September 2013 by European cardholders. This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172 % of all transactions.

Content

It contains only numerical input variables which are the result of a PCA transformation. Unfortunately, due to confidentiality issues, we cannot provide the original features and more background information about the data. Features V1, V2, ... V28 are the principal components obtained with PCA, the only features which have not been transformed with PCA are "Time" and "Amount". Feature "Time" contains the seconds elapsed between each transaction and the first transaction in the dataset. The feature "Amount" is the transaction Amount, this feature can be used for example-dependant cost-sensitive learning. Feature "Class" is the response variable and it takes value 1 in case of fraud and 0 otherwise.

Given the class imbalance ratio, we recommend measuring the accuracy using the Area Under the Precision-Recall Curve (AUPRC). Confusion matrix accuracy is not meaningful for unbalanced classification.

Capítulo 4

Course Structure & Outline

Hey everyone chris dutton here and welcome to thinking like an analyst this is a crah course desaigned for

4.1. Flavors of Analytics

All right let's take a minute and talk about the various roles or flavors of analytics because the thing is this field is very broad and it's very diverse so ir can be helpful to categorize various roles or job titles based on difertent types of skills so you may have seen venn diagrams out there that look something like this this version is adapted from learn.co we've got business intelligence skills they're in the blue bubble programming coding skills gray and math and stats skills in yellow now by visualizing skills in this way you can map various roles to the diagram based on the overlaps so for instance someone whi skews towards programmind and math might fall into the machine learning bucket people with and bi skills might fall into data engineering bi and math and stats maybe we call those people advanced analysts and those who use all three types of skills relatively equally might fall into the data science category and while this can be helpful to an extent remember that in reality this is fluid it's flexible and it's often somewhat subjective as well you know you can be a bi analyst who loves stats or programming or a machine learning engineer with exceptional business intelligence skills or literally any other combination of these categories the key is that this diagram this entire diagram represents the broader world of data analytics and well many types of roles fall under this analytics umbrella they're all aligned towards the same ultimate goal using data to make smart decisions. Now think about it that applies whether you're a bi analyst a statistician a data scientist or an everyday excel jockey and the difference is they come down to things one the types of problems that you're trying to solve and two the types of tools that you're using to solve them so in the next lesson we'll dive into one of the most common comparisons out there business intelligence versus data science but for now what's important to specific role you play we're all playing for the same team

Capítulo 5

Ingeniero de datos

Es un profesional que tiene pasión por la automatización, confianza en el análisis y los datos. Sus habilidades se enfocan en la extracción, transformación y carga de datos (ETL). Automatización de procesos de carga de datos. Procesamiento de datos de diversas fuentes y en cómputo paralelo. Tiene conocimientos en programación con Python y bases solidas de ingeniería de software, manejo de datos estructurados y no estructurados, cómputo, almacenamiento y bases de datos en la nube. ETL con herramientas como SQL, Apache Spark, Airflow, Hadoop, AWS, Google Cloud, Azure, entre otros.

5.1. Módulo 1

1. Ask:
2. Prepare:
3. Procesos:
4. Analyze:
5. Share:
6. Act:

5.2. Problemas con datos

Cuańdo se preparan los datos para un análisis, se pueden presentar problemas como:

- No tenemos los datos que necesitamos.
- Los datos que tenemos no son suficientes.
- Duplicados
- Valores faltantes
- Valores atípicos

- Datos incorrectos
- Datos inconsistentes
- Datos desactualizados
- Datos no estructurados
- Datos no normalizados
- Datos no limpios
- Datos no estandarizados
- Datos no validados
- Datos no verificados
- Datos no confiables
- Datos no seguros
- Datos no accesibles
- Datos no escalables
- Datos no flexibles
- Datos no eficientes
- Datos no eficaces

En muchos casos se suele usar lo que se conoce como datos indirectos en lugar de datos reales. En otros casos, puede no haber un sustituto razonable y la única opción será recopilar más datos.

5.2.1. Falta de datos

posibles soluciones:

1. Recopila datos en menor escala para realizar un análisis preliminar y, luego, pide más tiempo para completar el análisis después de haber recopilado más información.
2. Si no hay tiempo para recopilar información, realizar el análisis utilizando datos indirectos de otros conjuntos de datos. Está es la solución más común.

5.2.2. Muy pocos datos

Soluciones posibles:

1. Realiza el análisis utilizando datos indirectos junto con datos reales.
2. Ajusta tu análisis para alinearlo con los datos que ya tienes.

5.2.3. Datos incorrectos, incluidos los datos con errores

Soluciones posibles:

1. Si tienes datos incorrectos porque los requerimientoss no fueron bien interpretados, comunica los requerimientos nuevamente.

   Example 5.2.1.
   - Si necesitas datos sobre mujeres y votantes y recibiste datos sobre hombres votantes, vuelve a comunicar qué datos necesitas.
   - Si los datos se recopilaron incorrectamente, recopila los datos nuevamente.
   - Si los datos se ingresaron incorrectamente, corrige los datos en la fuente.
   - Si los datos se procesaron incorrectamente, procesa los datos nuevamente.
   - Si los datos se analizaron incorrectamente, analiza los datos nuevamente.
   - Si los datos se presentaron incorrectamente, presenta los datos nuevamente.

2. Identifica los errores en los datos y, cuando sea posible, corrígelos en la fuente buscando el patrón de errores.

   Example 5.2.2.

   Si tus datos se encuentran en una hoja de cálculo y hay una instrucción condicional o datos booleanos que generan errores en los cálculos, modifica la instrucción condicional en lugar de corregir los valores calculados.

3. Si no puedes corregir los errores en los datos tú mismo, puedes ignorarlos y seguir adelante con el análisis en caso de que el tamaño de tu muestra aún sea lo suficientemente grande como para poder ignorar esos datos y que eso no ocasione un sesgo sistemático.

   Example 5.2.3.

   Si tu conjunto de datos es una traducción de otro idioma y alguna traducción no tiene sentido, puedes ignorar los datos con traducciones erróneas y seguir adelante con el análisis de los otros datos.

4. Nota: a veces los datos con errores pueden ser una señal de advertencia sobre la falta de confiabilidad de los datos. Utiliza tu juicio.

Diagrama de toma de decisiones para recordar cómo manejar los errores en los datos o la falta de datos:

1. Data Errors:

5.3. El tamaño de la muestra

La población

Tamaño de la muestra(Sample size) Usar una parte de una población el objetivo es obtener suficiente información de un grupo pequeño de personas para hacer inferencias sobre toda

la población. dentro de una población para formular predicciones o conclusiones sobre la población total. La muestra asegura el grado respecto del cual puedes estar confiado en que tus conclusiones representan con presición a la población.

Nota: Cuando utilizas únicamente una muestra pequeña de una población, puede llevar a la incertidumbre en tus conclusiones. No puedes estar el 100 % seguro de que tus estadísticas son una representación precisa y completa de la población. Esto lleva a un sesgo del muestreo, ocurre cuando la muestra no es representativa de la población en su conjunto. Esto significa que algunos miembros de la población están siendo sobre o subrepresentados.

Existen métodos para solucionar el sesgo del muestreo.

1. Muestreo aleatorio

Cómo análista de datos es bueno saber que los datos que se van a analizar son representativos de una población y sirven para el objetivo.

Capítulo 6

Big-Data

Qué es Map Reduce?

Paradigma de programación que permite trabajar en ambientes distribuidos usando una gran cantidad de servidores que conforman un clúster.

1. **Map:** recibe un conjunto de datos y lo transforma en un segundo conjunto de datos cuyos elementos se presentan en tuplas (clave - valor)
2. **filter:**
3. **Reduce:** Utiliza como entrada las tuplas generadas por el map y genera un conjunto de datos de salida reducido a partir de la combinación de los datos recibidos

¿Cómo trabaja MapReduce?

Tiene tres etapas Map(), Shuffle y Reduce().

Qué es un RDD

Un RDD (Dataset distribuido Resiliente)

Transformaciones: Las transformaciones se aplican a RDD y devuelven un RDD.

1. **Map:**
2. **filter:**
3. **Reduce:**
4. **flatMap:**
5. **mapPartitions:**
6. **mapPartitionsWithIndex:**
7. **sample:**
8. **union:**

9. **intersection:**
10. **distinct:**
11. **groupByKey:**

Acciones:

1.

https://spark.apache.org/docs/latest/api/python/index.html

6.1. Roles

¿Quién define que variables que variables van a quedar en una fuente de datos que se este extrayendo?

La desición involucra a varios roles y depende del contexto del proyecto y objetivos del análisis. A continuación:

Data Engineer: Rol: Son responsables de la contrucción y mantenimiento de los pipelines de datos. Pueden decidir qué varaibles extraer en función de la viabilidad técnica y calidad de los datos.

Criterio de selección: Pueden excluir variables con baja calidad de datos (por ejemplo muchas ausencias o valores inconsistentes) o datos que no sean escalables para el almacenamiento y procesamiento.

Data Analyst: Rol: Son los responsables de examinar y comprender los datos para extraer información útil. A menudo, son quienes deciden qué variables son relevantes para el análisis.

Criterio de selección: Se basan en la relevancia de las variables para las preguntas de negocio o hipótesis que se desean validar. Pueden realizar un análisis exploratorio para identificar cuáles son los datos más significativos.

Data Scientist: Rol: Diseñan modelos predictivos o de aprendizaje automático y definen qué variables se usarán como características para estos modelos.

Criterio de selección: Aplican técnicas de selección de características (feature selection) para identificar las variables más importantes o realizar transformaciones para crear nuevas variables a partir de las existentes.

Usuario de Negocio(Stakeholders): Rol: Los usuarios de negocio, como gerentes o líderes de proyectos, son quienes tienen una visión clara de los objetivos del negocio y las decisiones que necesitan respaldarse con datos.

Criterio de selección: Pueden definir qué datos son necesarios para tomar decisiones estratégicas o monitorear métricas clave de desempleño (KPIs).

Arquitectos de Datos: Rol: Los arquitectos de datos son responsables del diseño de la infraestructura de datos. Pueden sugerir qué variables se deben extraer con base en la arquitectura existente y las limitaciones de almacenamiento o procesamiento.

Criterio de selección: Consideran la eficiencia de almacenamiento y procesamiento, así como la integración con otras fuentes de datos.

Gobernanza de Datos: Rol: Los equipos encargados de la gobernanza de datos aseguran que los datos utilizados cumplan con normativas y políticas de la organización.

Criterio de selección: Pueden excluir variables que contengan datos sensibles o que no cumplan con regulaciones de privacidad, como normativas GDPR o CCPA.

Proceso de Decisión Típico

1. Definir los Objetivos del Proyecto: Se establecen los objetivos del análisis o del proyecto de datos.
2. Selección Inicial de Variables: Basado en los requisitos de negocio y las hipótesis iniciales, se realiza una selección preliminar de variables.
3. Evaluación TÃ©cnica y de Calidad: Los ingenieros y analistas de datos evalÃºan la calidad y viabilidad de extraer y utilizar las variables seleccionadas.
4. Iteración y Refinamiento: Puede realizarse un proceso iterativo para ajustar la selección de variables con base en el análisis exploratorio de datos, la retroalimentación de los usuarios de negocio o los resultados de modelos preliminares.
5. Aprobación Final: Los stakeholders y los responsables de la gobernanza de datos aprueban la selección final.

La selección de variables es un proceso colaborativo que involucra a múltiples roles y se adapta a los objetivos específicos del proyecto. Â¿'Tienes un contexto particular en mente o algún proyecto del cual te gustaría discutir más a fondo?

De [8].

Parte II

Metodologías

Capítulo 7

CRISP-DM

CRISP-DM, que son las siglas de Cross-Industry Standard Process for Data Mining, es un método probado para orientar sus trabajos de minería de datos.

Parte III

Bases de Datos

Capítulo 8

Relaciones

Objetos

Entidad rectangulos

Capítulo 9

MySQL

9.1. Funciones de control de flujo

Case: Esta expresión nos permite realizar la condición y devolver el primer valor que cumpla con dicha condición

Example 9.1.1.
Primer ejemplo

```
select
case 1
when 1 then 'uno'
when 2 then 'dos'
else 'otro n\'umero'
end as valor;

segundo ejemplo:

select idFactura, idProducto,

case
when cantidad > 2 then 'M\'as de dos productos vendidos'
when cantidad = 2 then 'Dos productos vendidos'
else 'Menos de dos productos vendidos'
end as cantidad
from detalle_factura;

Tercer ejemplo

select nombre,
case
when email IS NULL then 'No tiene email registrado'
else 'email'
end as email,
pais
from cliente;
```

Podemos ver que es una sentencia muy similar a switch de vuelve el primer caso que cumpla la condición.

IF :

Example 9.1.2.

```
Primer ejemplo
select if(1 < 2, true, false) as resultado;

segundo ejemplo

select
idProducto,
if(cantidad > 1, cantidad*precioUnitario, precioUnitario) as total
from detalle_factura;

Tercer ejemplo

select
nombre,
if(fechaIngreso < '2016-12-31', concat(idEmpleado, '-16'),
if(fechaIngreso < '2017-12-31', concat(idEmpleado, '-17'),
if(fechaIngreso < '2018-12-31', concat(idEmpleado, '-18'),
concat(idEmpleado,'-19')
  )
   )
   ) as codigo
from empleado;
```

IFNULL y NULLIF: IFNULL nos permite evaluar una primera expresión, si esta expresión es null, entonces devolverá el segundo valor pasado por parámetro y NULLIF :

Example 9.1.3.

```
Primer ejemplo:

select ifnull(null, 'texto') as resultado;

Segundo ejemplo:

En este ejemplo devuelve los contactos de la tabla cliente en la
columna nombre si tiene email nos da el email pero si este campo
es null nos devuelve el tel\'efono

select nombre, ifnull(email, telefono) as contacto
from cliente;

Tercer ejemplo:

select nombre,
ifnull( (select email from cliente where idCliente = '14'),
'No tiene eamil registrado' )
```

```
as email
from cliente
where idCliente = '14';

select
nullif(
(select precioUnitario from producto where idProducto = 1),
(select )
)
```

NULLIF:

9.2. Subqueries

Es una declaración select en otra declaración, los subqueries devuelven datos de la consulta principal, los subqueries puede ser utilizados para agregar, actualizar, eliminar, enviar datos.

Example 9.2.1.

```
 Ejemplo n\'umero 1:
 Consiste en traer cuyos empleados tengan mayor salario al promedio:

 select idEmpleado, nombre, salario
 from empleado
 where salario > (select avg(salario) from empleado);

Ejemplo 2: Seleccionamos los empleados que no pertenezcan al departamento general:

select nombre, apelllido, idDepartamento
from empleado
where idDepartamento NOT IN (select idDepartamento
                             from departamento
                             where nombre like "%general%"
                              );
Ejemplo 3: facturas de los clientes que pertenezacan a Canada o Brasil:

select idCliente, idFactura
from factura
where idCliente IN( select idCliente
from cliente
where pais = 'canada' or pais = 'Brasil'
 );
```

Subconsultas:

Example 9.2.2.

```
select *
from factura
where idCliente = (select idCliente form cliente where nombre = 'Jordi');
```

```
select *
from producto
where precioUnitario <=
(select avg(precioUnitario) from producto where idCategoria = 5)
and idCategoria = 5;
```

comparando subconsultas

Subconsultas:

Example 9.2.3.

```
select idProducto, nombre
from producto
where idProducto = ANY (select idProducto from detalle_factura);
```

1. Introducción y fundamentos del aprendizaje automático.
    - Introducción y fundamentos del aprendizaje automático
    - Introducción, Definiciones, Sklearn Script básico de una simulación en ML
    - Regresión lineal y regresión logística + Taller
    - Taller con dataset grande limpieza de datos + train/test con métrica de scorebásica para regresión y para clasificación
2. Clasificación y selección de modelos
    - Paramétrico vs No paramétrico: K-nn vs Gaussian. Taller sobre los modelos, fronteras de decisión.
    - Selección de modelos, overfitting y regularización.
    - Taller con dataser real selección de modelos"k-fold, k-folds estratificado, k-fold por grupos, Boostrapping.
3. Árboles de decisión y máquinas de vectores de soporte
    - Árboles, Baggind + Random Forest.
    - Máquinas de Vectores de Soporte, One vs All, All vs All.
    - Tallér práctico comparación de modelos de la semana.
4. Boosting y selección de características
    - Boosting: AdaBoost, Gradient Boosting.
    - Selección de características e importancia de variables, PCA, LDA.
    - Taller de aplicación de las técnicas de la semana.

9.3. Unidad 1

9.3.1. Introducción y fundamentos del aprendizaje automático

Metodologías

dasd

1. KDD: Knowledge Discovery in Databases process.

   a) Selección: Selección e integración de los datos objetivo provenientes de fuentes multiples y heterogéneas.

   b) Procesamiento:

   - Eliminación de ruido y datos aislados o outliers.
   - Uso del conocimiento previo para eliminar las inconsistencias y los duplicados.
   - Escogencia y uso de estrategias para manejar la información faltante en los datasets.

   c) Transformación: Conversiń de los atributos

   - Preparación de los datos para el análisis.
   - Uso de transformaciones de atributos como: numerización, discretización, etc.
   - El resultado es conjunto de filas y columnas denominado vista minable.

   d) Minería de datos:

   - Análisis de los patrones o relaciones a descubrir.
   - Se comprende de 3 pasos:
     - Selección de la tarea.
     - Selección del algoritmo(s).
     - Aplicación/Entrenamiento del algoritmo.

   e) Implementación/Evaluación:

   - Implementación, interpretación o difución del modelo.

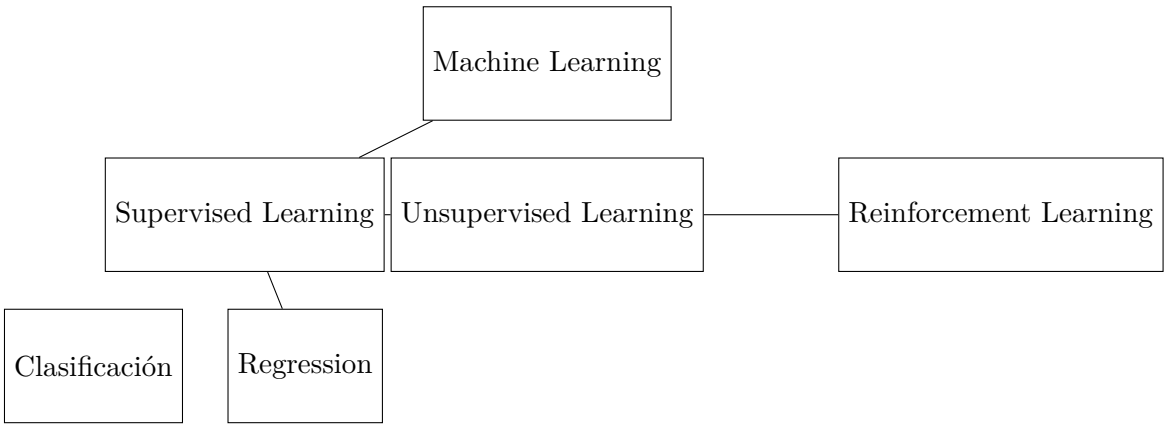
   f) Actualización y monitorización:

   - Consiste en ir revalidando el modelo con cierta frecencia sobre nuevos datos, con el objetivo de detectar si el modelo requiere una actualización.

2. CRISP DM

   - Entendimiento del negocio.
   - Entendimiento de los datos.
   - Preparación de los datos.
   - Modelado.
   - Evaluación.
   - Despliegue.

jlksjads


Machine Learning
Supervised Learning
Unsupervised Learning
Reinforcement Learning
Clasificación
Regression

Parte IV

Mathematics

Demostración: (del latín demostratio-onem) Prueba de una verdad por medio del raciocinio, partiendo de principios evidentes. En ciencia, en general se puede extender a la comprobación experimental de un principio o teoría.

Proposición (del latín propositio-onem): Enunciación de una verdad ya demostrada o que se ha de demostrar.

Silogismo (del latín syllogismus, y éste del griego): Argumento formado por tres proposiciones: premisa mayor, premisa menor y conclusión.

Enunciado (del latín enuntiatio-onem) Expresión o manifestación de una idea.

Hipótesis (del latín hypothesis, y éste del griego): Suposición de una idea, con el objetivo de deducir de ella alguna consecuencia.

Lema (del latín lemma, y éste del griego): Proposición cuya demostración antecede a un teorema.

Axioma (del latín axioma, y éste del griego): Principio, verdad, sentencia clara y evidente, que no necesita demostración.

Conclusión (del latín conclusio-onem): Proposición que se deduce o es consecuencia de las premisas.

Conjetura (del latín conjectura): Opinión fundada en probabilidades, indicios o apariencias.

Corolario (del latín corollarium) Proposición que por sí sola se deduce de lo ya demostrado.

Premisa (del latín praemissa): Cualquiera de las dos proposiciones lógicas de un silogismo, de donde se deduce la conclusión. En un silogismo, a la proposición mÃ¡s general se denomina mayor y a la otra menor.

Principio (del latín principium): Argumento considerado como origen, fundamentación o razón primera de un razonamiento.

Tesis (del latín thesis, y éste del griego): Conclusión o proposición que se mantiene con razonamientos.

Postulado (del latín postulatus): Proposición cuya verdad se admite sin pruebas y que es necesaria para servir de base en ulteriores razonamientos.

Propiedad (del latín propietas): Atributo o cualidad de una persona o cosa.

Teorema (del latín theorema, y éste del griego): Proposición que afirma una verdad susceptible de demostración.

Capítulo 10

Combinatorics

10.1. Four Basic Counting Principles

Definition 10.1.1. (Pairwise disjoint sets.) Let S be a set. A partition of S is a collection $S_1, S_2, \ldots, S_m$ of subsets of S such that each element of S is in exactly one of those subsets:

$$S_1 \cup S_2 \cup \cdots \cup S_m = S \quad \text{and} \quad S_i \cap S_j = \emptyset \quad \text{for all } i \neq j.$$

Thus, the sets $S_1, S_2, \ldots, S_m$ are pairwise disjoint sets, and their union is S. The subsets $S_1, S_2, \ldots, S_m$ are called the parts of the partition.

The numbers of objects of a set S is denoted by $|S|$ and is sometimes called the size of S.

Example 10.1.2.

1. $\Omega = \{T, H\}$ where T denotes tails and H denotes heads. The set Ω is a partition of the set of all possible outcomes of a single coin toss.

$$\Omega = \{T\} \cup \{H\}$$

10.1.1. Addition Principle

Definition 10.1.3. (Addition Principle.) Suppose that a set S is partitioned into pairwise disjoint parts $S_1, S_2, \ldots, S_n$. The number the of objects in S can be determined by finding the number of objects un each of the parts, and adding the number so obtained:

$$|S| = |S_1| + |S_2| + \ldots + |S_n|. \tag{10.1}$$

If the sets $S_1, S_2, \ldots, S_m$ are allowed to overlap, then a more profound principle, the inclusion-exclusion principle, is needed.

Example 10.1.4.

1.

10.1.2. Multiplication Principle

Definition 10.1.5. (Multiplication Principle.) Let S be a set of ordered pairs (a, b) of objects, where the first object a comer from a set of size p , and for each choice of object a there are q choices for object b. Then the size of S is $p \times q$:

$$|S| = p \times q. \tag{10.2}$$

The mulplication principle is actually a consequence of the addition principle.

A second useful formulation of the multiplication principle is as follows: If a first task has p outcomes and, no matter what the outcome of the first task, a second task has q outcomes, then the two tasks performed consecutively have $p \times q$ outcomes.

10.1.3. Subtration Principle

Definition 10.1.6. Subtraction Principle Let A be a set and let U be a larger ser containing A. Let

$$\overline{A} = U \setminus A = \{x \in U | x \notin A\}$$

be the complement of A in U. Then the number $|A|$ of objects in A is given by the rule

$$|A| = |U| - |\overline{A}|.$$

In applying the subtraction principle, the set U is usually some natural set sisting of all the objects under discussion(the so-called universal set). Using the subtraction principle makes sense only if it is easier to count the number of objects in U and in $\overline{A}$ than to count the number og objects in A.

10.1.4. Division Principle

Definition 10.1.7. Division Principle Let S be a finite set that is partitioned into k parts in such a way that each part contains the same number of objects. Then the number of parts in the partition is given by the rule

$$k = \frac{|S|}{\text{number of objects in a part}}.$$

Thus, we can determine the number of parts if we know the number of objects in S and the common value of the number of objects in the parts.

10.1.5. PERMUTATIONS OF SETS

Let r be a positive integer, by an r-permutation of a set S of n elements, we understand an ordered arrangement of r of the n elements.

We denote by $P(n, r)$ the number of r-permutations of an n-element set.

Theorem 10.1.8. *For n and r positive integers with $r \leq n$,*

$$P(n, r) = n(n-1)(n-2)\cdots(n-r+1) = \frac{n!}{(n-r)!}. \tag{10.3}$$

Demostración. $P(n,r) = \_\ \_\ \underbrace{\_\ \ldots\ \_}_{r-times}\ \_\ \_$ □

10.1.6. Principio de la adición

Si un suceso S_1 ocurre de a maneras y un suceso S_2 ocurre de b maneras entonces puede decirse que el número total de maneras en que puede ocurrir S_1 o el suceso S_2 es $a + b$.

Example 10.1.9.

1 **Ejemplo 1:** ¿Cuántos resultados diferentes se pueden obtener lanzando un dado o una moneda? Indetificamos que el experimento consiste en elegir que objeto lanzar, suponga:

$$S_1 : = \text{ lanzar el dado } \text{ y } S_2 : = \text{ lanzar la moneda.}$$

Si se elije el dado, entonces el número de resultados posibles es $\{1, 2, 3, 4, 5, 6\}$ o si elige la moneda el número de resultados posibles es $\{C, S\}$. Así el número total de resultados posibles para el suceso S_1 o S_2 es:

$$6 + 2 = 8$$

.

2 **Ejemplo 2:** ¿Cuántos resultados diferentes se pueden obtener lanzando un dado y una moneda? Considerando los mismo sucesos S_1 y S_2 del ejemplo anterior, el experimiento consiste en lanzar un dado y seguidamente lanzar una moneda. nuestro conjunto de resultados sería

$$\{(1, C), (1, S), (2, C), (2, S), \ldots, (6, C), (6, S)\}.$$

Así la cantidad de sucesos es $6 \cdot 2 = 12$.

10.1.7. Principio de la multiplicación

Definition 10.1.10. (Principio de la multiplicación) Si un experimento consiste de n etapas o sucesos y si la primera etapa o suceso puede ser realizada de n_1 maneras, la segunda etapa de n_2 maneras, la k-ésima etapa de n_k maneras, entonces el experimento completo puede ser realizado de $n_1 \cdot n_2 \cdot \ldots \cdot n_k$ maneras.

10.1.8. Permutaciones

¿De cuántas formas se pueden ordenar u organizar los elementos de un conjunto o de un subconjunto? Para responder tal pregunta, realicemos las siguientes definiciones.

Definition 10.1.11.

1. Ordenar: significa poner una cantidad de elementos en un orden específico que obedece a una regla lógica.
2. Organizar: implica poner una cantidad de elementos de forma que no tienen un orden específico o dicho orden no es importante dentro de un determinado contexto.

3. Conjunto:

4. Subconjunto:

5. Orden: corresponde a la aplicación de un criterio específico que le confiere cierta jerarquía a los elementos que forman parte de un conjunto.

Definimos lso conceptos de permutación y combinación:

Definition 10.1.12. Permuatción

Parte V

Análisis de series temporales

Capítulo 11

Introducción a las series temporales

Para estudiar está parte del libro nos fijaremos en el libro [7], para estudiar las series temporales.

Definition 11.0.1. ([7]) Una serie temporal es el resultado de observar los valores de una variable a lo largo del tiempo en intervalos regulares (cada día, cada mes, cada año, etc.)

Series estables en el tiempo o estacionarias.

1. Oscilan alredor de un nivel constante.
2. Su gráfico no muestra ninguna tendencia clara a crecer o decrecer en el tiempo.

Series no estables en el tiempo o no estacionarias.

1. La tendencia en el tiempo es variable, en sentido creciente o decreciente.
2. Presentan una tendencia evolutiva o cambiante en el tiempo.

Series estacionales

1. El valor medio de la variable observada dependa del intervalo seleccionado.

Cuándo el nivel de la serie no es estable decimos que la serie no es estacionaria.

Diccionario

Fluctuaciones: se refieren a los cambios, variaciones o desviaciones que experimenta una cantidad o un valor a lo largo del tiempo, de forma que no sigue un patrón completamente predecible. En otras palabras, las fluctuaciones son las oscilaciones o movimientos alrededor de un nivel o valor promedio, ya sea en una serie temporal o en cualquier sistema dinámico.

1. Naturaleza impredecible: Las fluctuaciones pueden ocurrir de manera aleatoria o debido a factores externos que influyen en el sistema.
2. Magnitud variable: Las fluctuaciones pueden ser pequeñas y casi imperceptibles, o pueden ser grandes y representar cambios significativos en los valores de la serie.
3. Dirección: Las fluctuaciones pueden ser hacia arriba o hacia abajo, es decir, pueden reflejar aumentos o disminuciones en el valor observado.

Tipos de fluctuaciones:

1. Fluctuaciones aleatorias(ruido): Son cambios que ocurren debido a factores impredecibles y no siguen ningún patrón específico. Se consideran ruido en los datos.
2. Fluctuaciones cíclicas o estacionales: En algunos sistemas, las fluctuaciones pueden seguir patrones repetitivos y predecibles, como los ciclos estacionales (verano/invierno), ciclos económicos, o patrones de demanda.

Efectos externos o efectos de intervención:

Bibliografía

[1] C.B. Liliana, A. Arunachalam, D. Selvamuthu *Introduction to probability and stochastic processes with applications*, 1ª ed., New Jersey: John Wiley & Sons, Inc, 2012.

[2] R.A. Brualdi. *Introductory Combinatorics*, 5ª ed., Republic of China: Prentice-Hall, 2009.

[3] "Uso de Jupyter Notebook en un entorno virtual", 2022. [En linea]. Disponible en: `https://es.acervolima.com/uso-de-jupyter-notebook-en-un-entorno-virtual/`. [Accedido: 21-jun-2022]

[4] "Create a Next.js App", 2022.[En linea]. Disponible en: `https://nextjs.org/learn/basics/create-nextjs-app`. [Accedido: 15-jun-2022]

[5] M.P. Do Carmo *Differential Geometry Of Curves & Surfaces*, 1ª ed., USA: Prentice-Hall, 1976.

[6] L. Leithold, , 2007. Editorial Harla. Méjico. *Cálculo con Geometría Analítica.*, 1ª ed., USA: Prentice-Hall, 1976.

[7] D. Peña Sánchez de Riviera *Análisis de series temporales.* 2ª ed. Madrid, España: Alianza Editorial, 2010.

[8] OpenAI, *ChatGPT.* Available: https://chat.openai.com/, [Accessed: 15-Oct-2024].