



UNIVERSIDAD
NACIONAL
DE COLOMBIA

CLASIFICACIÓN Y RECONOCIMIENTO DE PATRONES

JOHN W. BRANCH

Profesor Titular

Departamento de Ciencias de la Computación y de la Decisión

Director del Grupo de I+D en Inteligencia Artificial – GIDIA

jwbranch@unal.edu.co

<https://github.com/jwbranch/ClasificacionYReconocimientoDePatrones>

<https://www.coursera.org/programs/unal-iuukt>

METODOLOGÍA ENSEÑANZA – APRENDIZAJE

El aprendizaje sincrónico involucra estudios online a través de una plataforma. Este tipo de aprendizaje sólo ocurre en línea. Al estar en línea, el estudiante se mantiene en contacto con el docente y con sus compañeros. Se llama aprendizaje sincrónico porque la plataforma permite que los estudiantes pregunten al docente o compañeros de manera instantánea a través de herramientas como el chat o el video chat.

Sesiones Remotas vía Google.Meet Sincrónicas y Asincrónicas

El aprendizaje asincrónico puede ser llevado a cabo online u offline. El aprendizaje asincrónico implica un trabajo de curso proporcionado a través de la plataforma o el correo electrónico para que el estudiante desarrolle, de acuerdo a las orientaciones del docente, de forma independiente. Un beneficio que tiene el aprendizaje asincrónico es que el estudiante puede ir a su propio ritmo.

Descripción del Curso

La descripción, la agrupación, la clasificación y el reconocimiento, de patrones de forma automática, son problemas importantes en una gran variedad de aplicaciones de ingeniería, psicología, medicina, economía, biología, etc. El problema consiste en asignar automáticamente a una clase una muestra según las mediciones realizadas sobre la muestra. En el curso se estudiará la teoría necesaria para resolver este problema, y se aplicará la teoría en aplicaciones de diferentes áreas.

OBJETIVOS DEL CURSO

Identificar los componentes principales de un sistema de Clasificación y Reconocimiento de Patrones, a partir de **información visual**.

Estudiar los conceptos básicos de extracción de características, selección de características, clasificación y evaluación de desempeño

Implementar los algoritmos asociados con las distintas etapas y modelos de un sistema de Clasificación y Reconocimiento de Patrones.

Aplicar técnicas para establecer y comparar el desempeño de los clasificadores.

Aplicar métodos de Clasificación y Reconocimiento de Patrones para resolver un problema real.

Contenido

1. Introducción.
2. **Adquisición, Procesamiento y Etiquetado de Datos**
3. Extracción de Características.
4. Selección de Características.
5. Clasificación Supervisada.
6. Clasificación No Supervisada
7. Evaluación de Desempeño.
8. Aplicaciones y Casos de Éxito.

Bibliografía Recomendada

1. **Bishop, C.** Pattern Recognition and machine Learning, Springer, 2006.
2. **Bishop, C.** Neural Network for Pattern Recognition, New York, Oxford University Press Inc., Reprinted, 2005.
3. **Da Fontoura, L.; Marcondes, R.** Shape Analysis and Classification, Boca Raton, CRC Press, 2001.
4. **Duda, R.; Hart, P.; Stork, D.** Pattern Classification, New York, John Wiley & Sons, Inc., 2001.
5. **Hastie, T.; Tibshirani, R.; Friedman, J.** The Elements of Statistical Learning, Springer, 2001.
6. **Marsland, S.** Machine Learning: An algorithmic Perspective, CRC Press, 2009.
7. **Mery, D.** X-ray Testing for Computer Vision, Springer, 2015.
8. Nixon, M.; Aguado, A. Feature Extraction & Image Processing, Amsterdam, Elsevier, 2004.
9. **Tou, J. T.; Gonzalez, R. C.** Pattern Recognition Principles, London-Amsterdam-Dom Mills, Ontario-Sydney-Tokyo. Addison-Wesley Publishing Company, 1974.
10. **Webb, A.** Statistical Pattern Recognition, Wiley, Second Edition, 2002.
11. **Witten, I.H; Frank, E.** **Data Mining:** Practical Machine Learning Tools and Techniques, Elsevier, Second Edition, 2005.
12. **Artículos seleccionados de IEEE** Transaction on Pattern Analysis and Machine Intelligence y de Proceedings of International Conferences on Pattern Recognition and Computer Vision.
13. **Scikit-learn** :Biblioteca de Aprendizaje Automático de software gratuito para el lenguaje de programación Python <https://scikit-learn.org/stable/>
14. **Kaggle:** Is an online community of data scientists and machine learners, owned by Google LLC. Kaggle allows users to find and publish data sets, explore <https://www.kaggle.com/>

EVALUACIÓN



Certificado Coursera <i>Sesenta años de inteligencia artificial – UNAM (Obligatorio)</i> <i>IA para todos – Andrew Ng (Obligatorio)</i> <i>Structuring Machine Learning Projects – Andrew Ng (Obligatorio)</i> <i>Detección de Objetos – UAB (Obligatorio)</i>	20% (Máx. 31 de Dic/2020)
Informe de Lectura (Individual) <i>From ImageNet to Image Classification-Contextualizing Progress on Benchmarks.</i> <i>Pattern classification with missing data - A Review.</i> <i>Statistical Pattern Recognition - A Review.</i> <i>Towards Automatic Visual Inspection.</i>	20% (Máx. 15 de Dic/2020)
Trabajo Final (Debe ser en Grupo 3 ó 5 personas) Obtener el conjunto de datos (Imágenes) de los siguientes repositorios o cualquier otro disponible: http://www.ics.uci.edu/~mlearn/databases/ https://www.kaggle.com/datasets Proponer una solución para un caso de estudio, comparando mínimo tres (3) técnicas de CRP.	60% (Máx. 2 de feb/2021)

CLASIFICACIÓN Y RECONOCIMIENTO DE PATRONES

Adquisición, Procesamiento y Etiquetado de Datos

JOHN W. BRANCH

Profesor Titular

Departamento de Ciencias de la Computación y de la Decisión

Director del Grupo de I+D en Inteligencia Artificial – GIDIA

jwbranch@unal.edu.co

<https://github.com/jwbranch/ClasificacionYReconocimientoDePatrones>

<https://www.coursera.org/programs/unal-iuukt>

Datos



En el mundo, la cantidad de datos recolectados cada segundo a través de diferentes dispositivos electrónicos es realmente grande. Con un conjunto de estos, y utilizando técnicas matemáticas y computacionales, somos capaces de crear información.

Existen diferentes métodos y caminos para convertir los datos en información. Algunos buscan entender los datos desde el área estadística, mientras otros buscan predecir ciertos valores (clasificación y regresión).

Datos

En la actualidad, la recolección de datos e información se ha vuelto uno de los objetivos principales en la mayoría de empresas. Con ellos, una compañía puede mejorar sus estrategias de ventas, así como atraer a nuevos clientes dependiendo de sus gustos y/o necesidades.



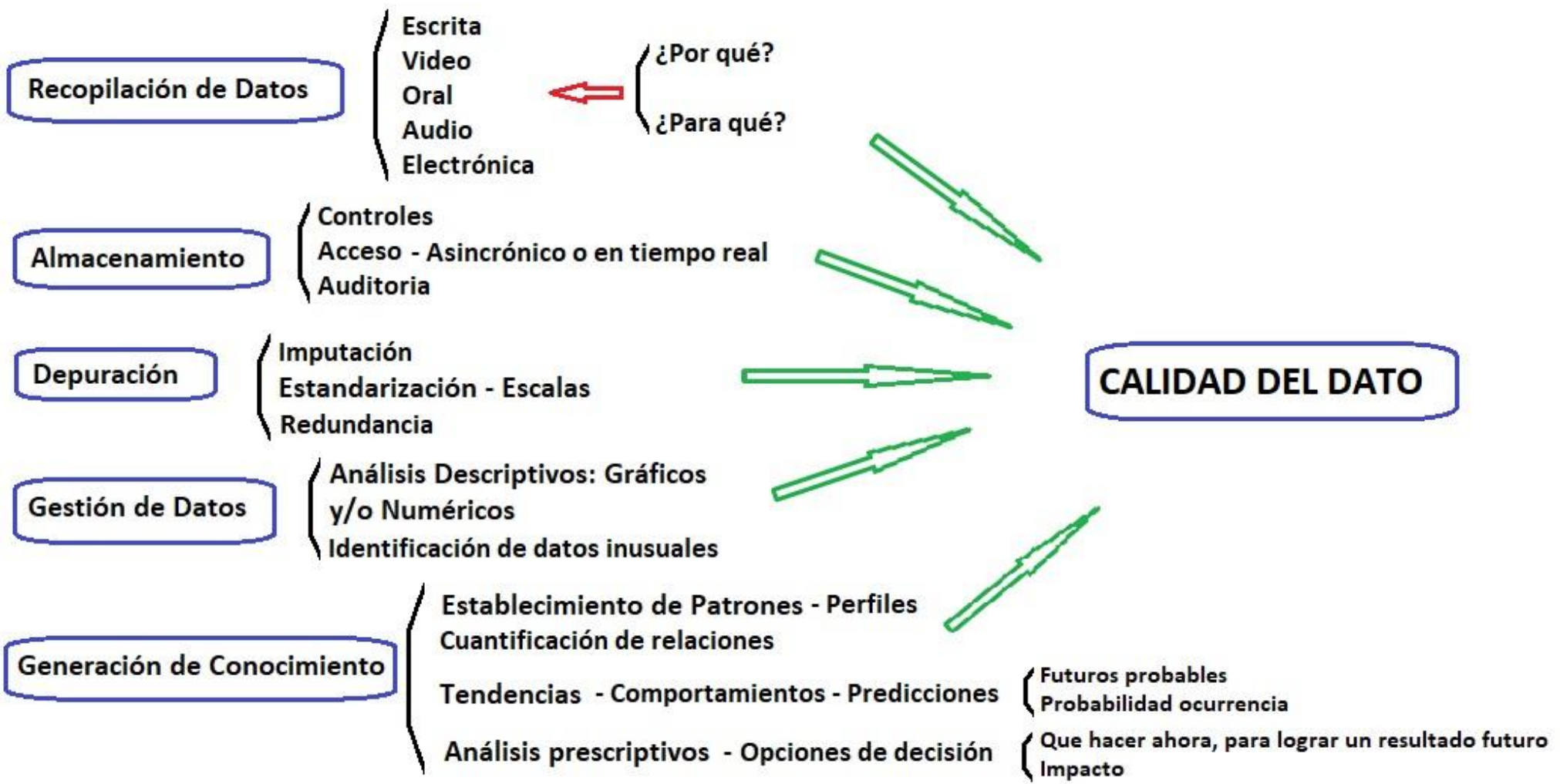
Datos

Actualmente, la utilización de datos no solo nos sirve para un análisis estadístico. Con los avances computacionales en inteligencia artificial, hemos logrado poder automatizar procesos que, años atrás, nos era poco rentable.

Utilizando diferentes técnicas, se han remplazado procesos muy costosos en tiempo y dinero por sistemas automáticos con el mismo o mejor desempeño obtenido por un grupo de profesionales.



GESTIÓN DE INFORMACIÓN





Adquisición de Datos

Los datos pueden ser clasificados en 4 dominios, dependiendo de su origen:

TEXTO

CUANDO **EL** OJO **VE** UN **COLOR** SE **EXCITA** INMEDIATAMENTE, Y ÉSTA **ES** SU **NATURALEZA**, ESPONTÁNEA Y DE **NECESIDAD**, PRODUCIR **OTRA** EN LA QUE **EL** COLOR **ORIGINAL** COMPRENDE LA ESCALA **CROMÁTICA** ENTERA. **UN** ÚNICO **COLOR** EXCITA, **MEDIANTE** UNA **SENSACIÓN** ESPECÍFICA, LA TENDENCIA A LA **UNIVERSALIDAD**. EN **ESTO** RESIDE LA LEY **FUNDAMENTAL** DE **TODA** ARMONÍA **DE** LOS **COLORES...**

AUDIO



VIDEO



IMAGEN



Datos - Imagen

Natural Images - Kaggle

Clasificación de 8 diferentes clases de un conjunto de imágenes “naturales”.



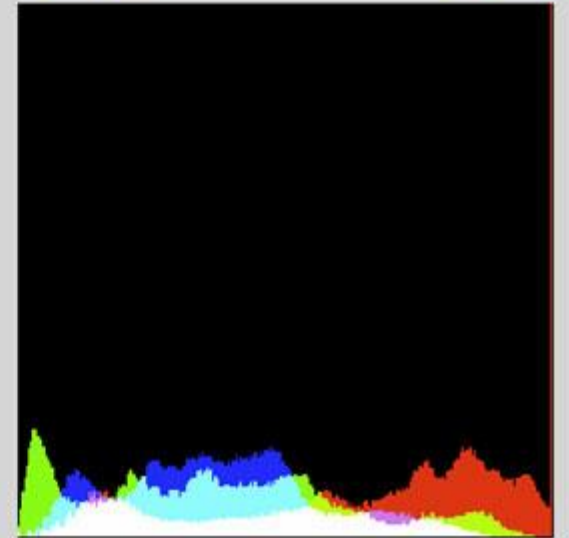
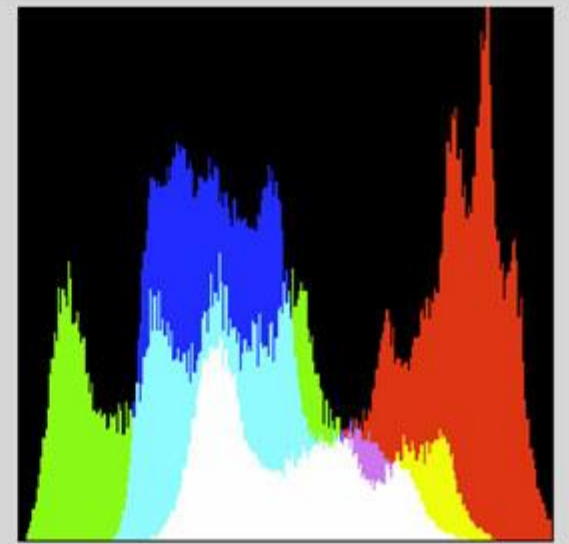
Preprocesamiento de Datos

Los datos a utilizar deben pasar por un proceso de preprocesamiento. Esto para seguir un estándar en los datos y lograr un mayor desempeño y exactitud a la hora de resolver el problema.

Si los datos no pasan por este proceso, los resultados en las futuras etapas no podrán alcanzar los valores reales de precisión posibles.



Preprocesamiento de Datos - Imagen



Etiquetado

El etiquetado de datos es una de las tareas más importantes a la hora de extraer información de estos.

Al tener imágenes etiquetadas correctamente, el computador es capaz de aprender a diferenciar entre diferentes clases. Por ejemplo, entre gatos y perros

Machine Learning:

Sample



Label



dog



cat



horse

Human Learning:

We learn through



Examples

Long Ear Black nose



dog

Diagrams

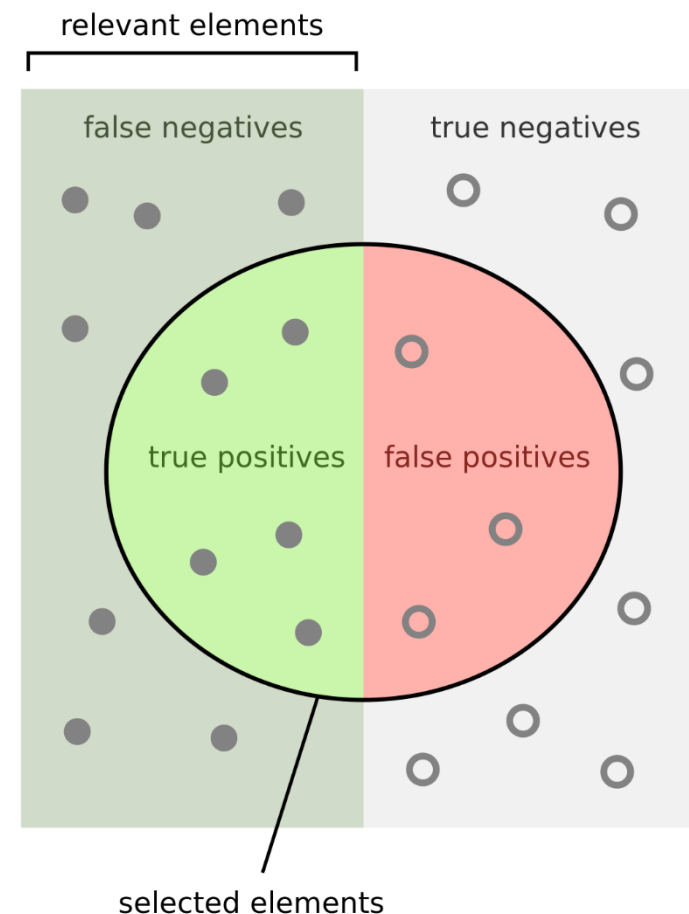


Comparisons

Etiquetado

En imágenes, los datos pre-procesados y útiles (para el problema de clasificación) deben estar compuestos por un ejemplo (la imagen de un tamaño y rango de color específico) y una etiqueta (la clase a la que pertenece tal ejemplo).

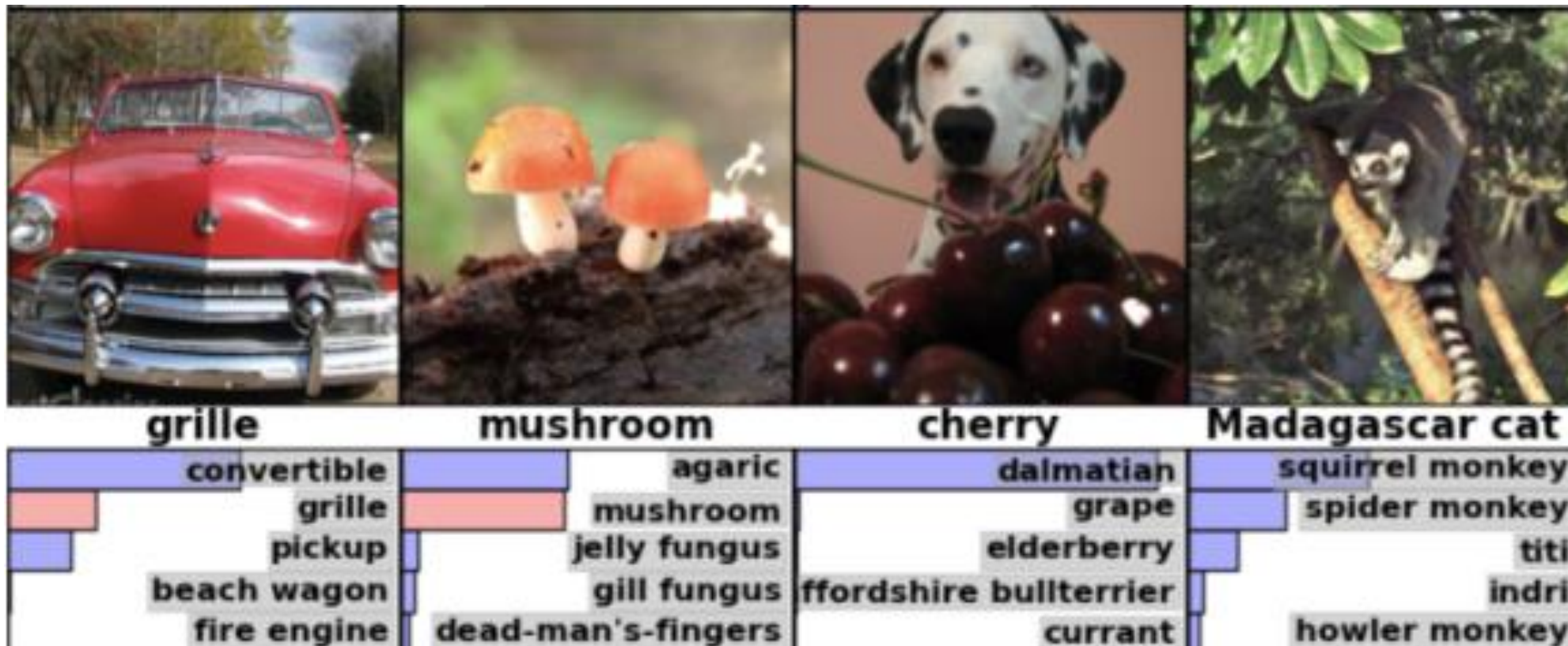
Una etiqueta errada de los ejemplos nos llevaría a tener dato que impediría a nuestro modelo alcanzar su máximo potencial.



Etiquetado



Etiquetado



Etiquetado

En el proceso de clasificación de imágenes es importante que nuestros datos etiquetados cumplan una distribución igual o parecida entre ellos.

Es decir, que nuestras clases tengan el mismo número de ejemplos con la misma desviación estándar en la información de las imágenes.

Image classification

Easiest classes

red fox (100) hen-of-the-woods (100) ibex (100) goldfinch (100) flat-coated retriever (100)



tiger (100)



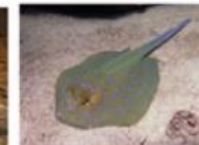
hamster (100)



porcupine (100)



stingray (100)



Blenheim spaniel (100)



Hardest classes

muzzle (71) hatchet (68) water bottle (68) velvet (68) loupe (66)



hook (66)



spotlight (66)



ladle (65)



restaurant (64)

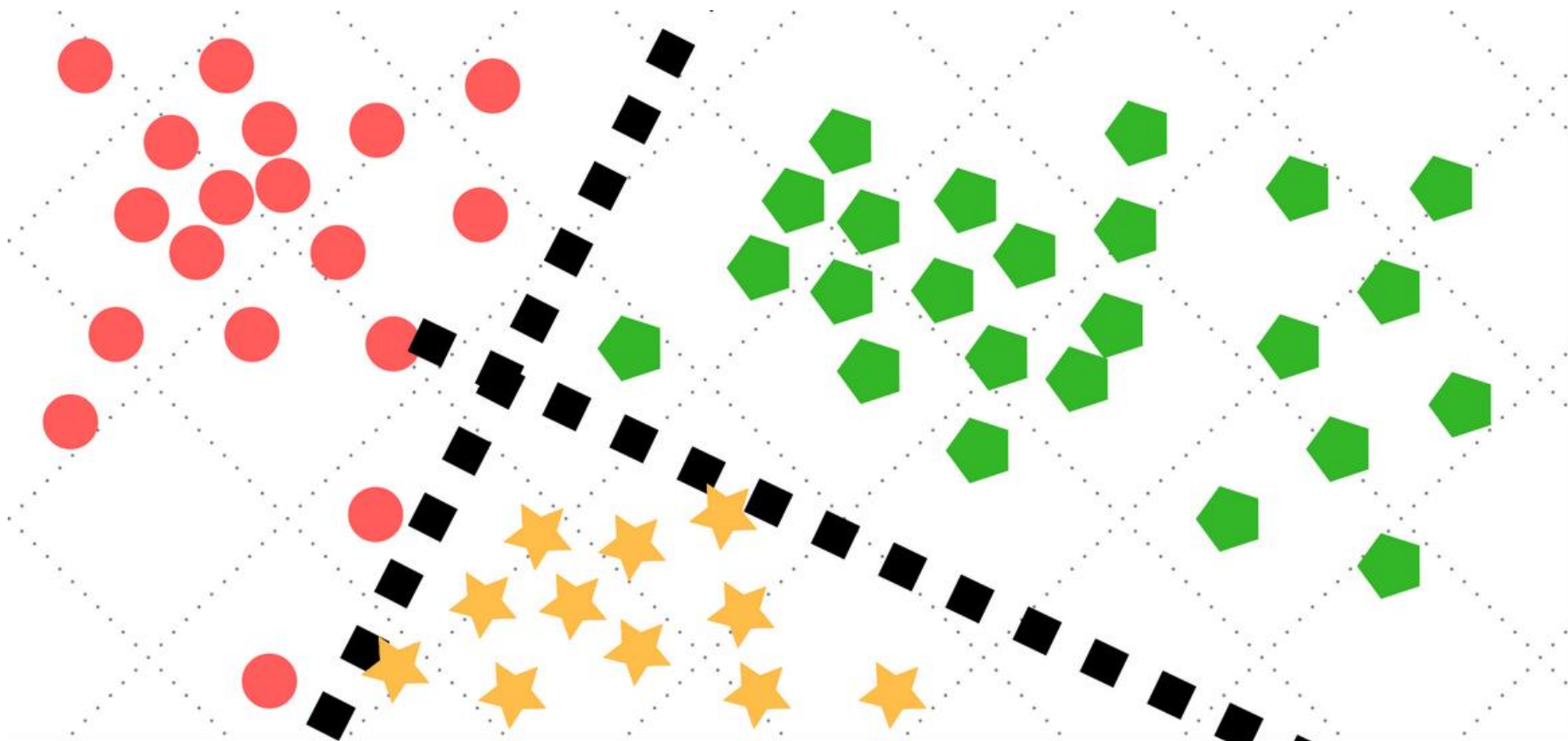


letter opener (59)



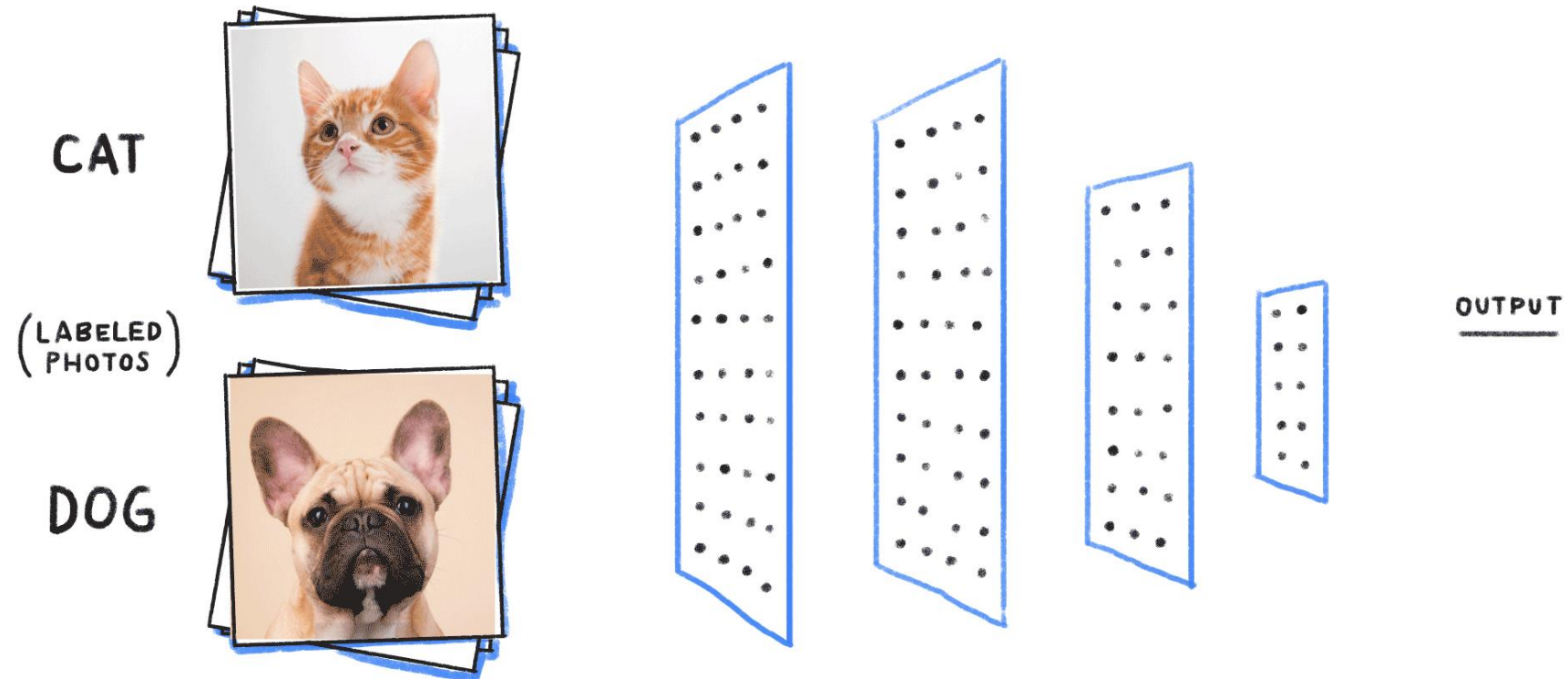
Etiquetado

Todo esto, con el fin de realizar tareas como:



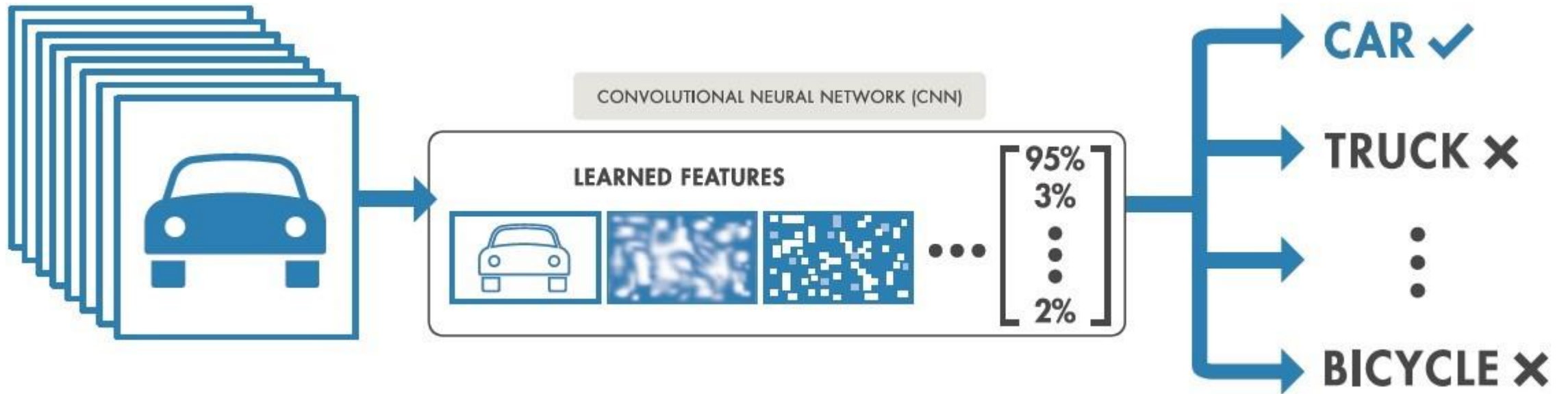
Ejemplos

Clasificación y Reconocimiento de Patrones



Ejemplos

Clasificación y Reconocimiento de Patrones

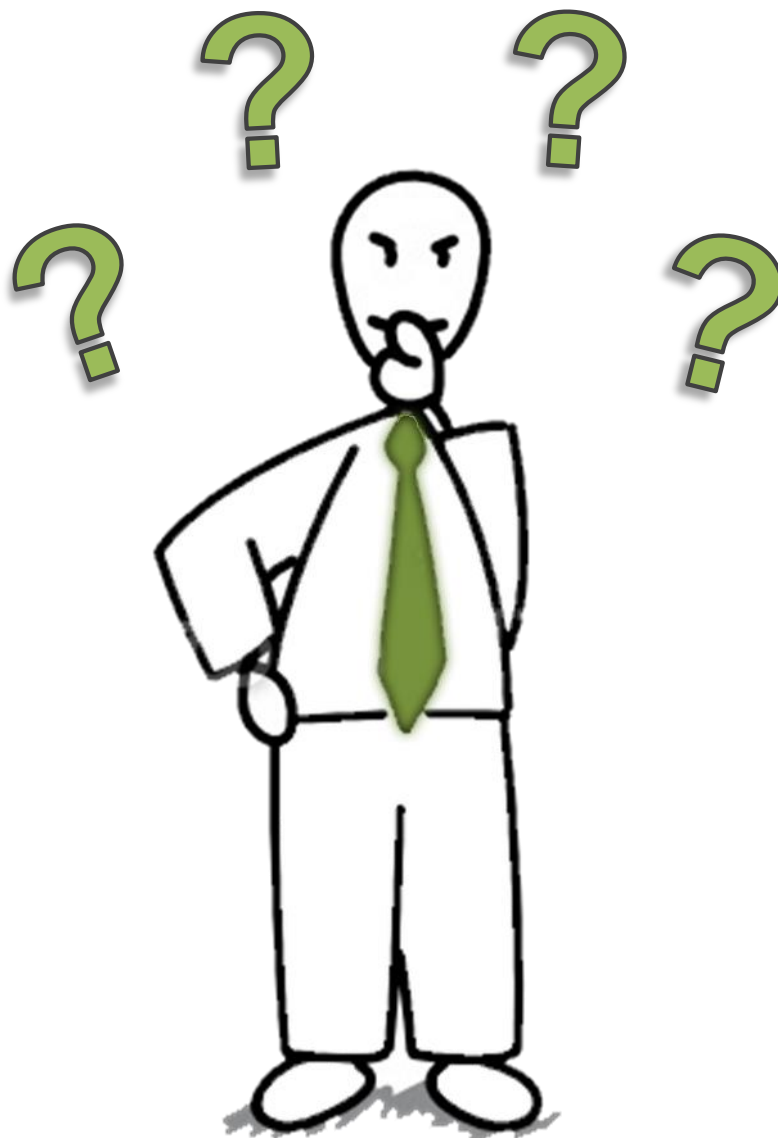


Etiquetado

Si el proceso de etiquetado no se realizó correctamente, este puede traer grandes consecuencias en el resultado final del modelo diseñado.



Preguntas



MOTIVACIÓN

🦋 OBSERVE EL VIDEO Y RESPONDA A LAS SIGUIENTES PREGUNTAS:

🦋 ¿CUÁNTOS DATOS SE REQUIEREN PARA ENTRENAR UN SISTEMA DE VISIÓN ARTIFICIAL?

🦋 ¿ES POSIBLE DECIR QUE LOS COMPUTADORES YA SOBREPASARON LA CAPACIDAD HUMANA?

🦋 ¿QUÉ PROBLEMAS EVIDENCIAN LOS SISTEMAS DE VISIÓN ARTIFICIAL, Y EN GENERAL DE LOS SISTEMAS DE RECONOCIMIENTO DE PATRONES?



<https://www.ted.com/talks/fei-fei-li-how-we-re-teaching-computers-to-understand-pictures?language=es>



UNIVERSIDAD
NACIONAL
DE COLOMBIA