



Aprende Machine Learning

antes de que sea demasiado tarde



7 PASOS
MACHINE
LEARNING



GENERAL

7 pasos del Machine Learning para construir tu máquina

🕒 septiembre 11, 2017 by Na8

Describiré los 7 pasos genéricos que debes seguir para construir tu propia Inteligencia Artificial con [Machine Learning](#).

Paso 1: Colectar Datos

Dada la problemática que deseas resolver, deberás investigar y obtener datos que utilizaras para *alimentar a tu máquina*. Importa mucho la calidad y cantidad de información que consigas ya que **impactará directamente en lo bien o mal** que luego funcione nuestro modelo. Puede que tengas la información en una base de datos ya existente o que la debas crear desde cero. Si es un pequeño proyecto puedes crear una planilla de cálculos que luego se exportará fácilmente como archivo csv. También es frecuente utilizar la técnica de [web scraping](#) para recopilar información de manera automática de diversas fuentes (y/o servicios rest/ APIs).

NO olvides hacer el EDA: [Análisis Exploratorio de Datos](#)

Paso 2: Preparar los datos

Es importante mezclar «las cartas» que obtengas ya que **el orden en que se procesen los datos dentro de tu máquina no debe de ser determinante**.

También es un buen momento para hacer visualizaciones de nuestros datos y revisar si hay correlaciones entre las distintas características («features», suelen ser las columnas de nuestra base datos o archivo) que obtuvimos. Habrá que hacer Selección de Características, pues las que elijamos impactarán directamente en los tiempos de ejecución y en los resultados, también podremos hacer **reducción de dimensiones aplicando PCA** si fuera necesario. **Deberemos tener balanceada la cantidad de datos** que tenemos para cada resultado(clase), para que sea representativo, ya que si no, el aprendizaje podrá ser tendencioso hacia un tipo de respuesta y cuando nuestro modelo intente generalizar el conocimiento fallará.

También **deberemos separar los datos en en dos grupos: uno para entrenamiento y otro para evaluación del modelo**. Podemos fraccionar aproximadamente en una proporción de 80/20 pero puede variar según el caso y el volumen de datos que tengamos.

En esta etapa también podemos preprocesar nuestros datos normalizando, eliminar duplicados y hacer corrección de errores.

Paso 3: Elegir el modelo

Existen diversos modelos que podemos elegir de acuerdo al objetivo que tengamos: utilizaremos algoritmos de **clasificación**, **predicción**, **regresión lineal**, **clustering (ejemplo k-means ó k-nearest neighbor)**, **Deep Learning** (ej: **red neuronal**), **bayesiano**, etc y podrá haber variantes si lo que vamos a procesar son imágenes, sonido, texto, valores numéricos. En la siguiente tabla veremos algunos modelos y sus aplicaciones

Modelo	Aplicaciones (Ejemplo de uso)
Logistic Regression	Predicción de precios de inmuebles
Fully connected networks	Clasificación
Convolutional Neural Networks	Procesamiento de imágenes para poder encontrar gatitos en las fotos
Recurrent Neural Networks	Reconocimiento de Voz
Random Forest	Detección de Fraude
Reinforcement Learning	Enseñarle a la máquina a jugar videojuegos y vencer!
Generative Models	Creación de imágenes
K-means	Crear Clusters a partir de datos sin etiquetar. Segmentar audiencias o Inventarios
k-Nearest Neighbors	motores de recomendación (por similitud/cercanía)

Aquí puedes ver más [aplicaciones del ML y la diferencia entre aprendizaje supervisado y no supervisado](#)

Paso 4 Entrenar nuestra máquina

Utilizaremos el set de datos de entrenamiento para ejecutar nuestra máquina y deberemos de ver una mejora incremental (para la predicción). Recordar inicializar los «pesos» de nuestro modelo aleatoriamente, **los pesos son los valores que multiplican o afectan a las relaciones entre las entradas y las salidas**, se irán ajustando automáticamente por el algoritmo seleccionado cuanto más se entrena. Revisar los resultados obtenidos y corregir (por ej. inclinación de la pendiente) y volver a iterar...

Paso 5: Evaluación

Deberemos **comprobar la máquina creada contra nuestro set de datos de Evaluación** que *contiene entradas que el modelo desconoce* y verificar la precisión de nuestro modelo ya entrenado. Si la exactitud es menor o igual al 50% ese modelo no será útil ya que sería como lanzar una moneda al aire para tomar decisiones. Si alcanzamos un 90% o más podremos tener una buena confianza en los resultados que nos otorga el modelo.

Paso 6: Parameter Tuning (configuración de parámetros)

Si durante la evaluación no obtuvimos buenas predicciones y nuestra precisión no es la mínima deseada es posible que tengamos **problemas de overfitting (ó underfitting)** y deberemos retornar al paso de entrenamiento (4) haciendo antes *una nueva configuración de parámetros de nuestro modelo*. Podemos incrementar la cantidad de veces que iteramos nuestros datos de entrenamiento (EPOCHs). Otro parámetro importante es el conocido como «Learning Rate» (taza de aprendizaje) que suele ser un valor que multiplica al gradiente para acercarlo poco a poco al mínimo global (o local) para minimizar el coste de la función. No es lo mismo incrementar nuestros valores en 0,1 unidades que de 0,001 esto puede afectar significativamente el tiempo de ejecución del modelo. También se puede indicar el máximo error permitido de nuestro modelo. Podemos pasar de tardar unos minutos a horas (y días) en entrenar nuestra máquina. A estos parámetros muchas veces se les llama Hiperparámetros. **Este «tuneo» sigue siendo más un arte que una ciencia y se ira mejorando a medida que experimentamos.** Suele haber muchos parámetros para ir ajustando y al combinarlos se pueden disparar todas nuestras opciones. Cada algoritmo tiene sus propios parámetros a ajustar. Por nombrar alguno más, en las **Redes**

Neuronales Artificiales deberemos definir en su arquitectura la cantidad de hidden layers que tendrá e ir probando con más o con menos y **con cuantas neuronas cada capa**. Este será un trabajo de gran esfuerzo y paciencia para dar con buenos resultados.

Paso 7 : Predicción o Inferencia

Ya estamos listos para utilizar nuestro modelo de Aprendizaje Automático! con nueva información y comenzar a predecir o inferir resultados «en la vida real»... Imaginemos que si este paso fuera pasar de un simulador de un coche a **aplicar nuestra máquina en un coche de verdad**... es un gran momento y también será un emocionante desafío poner a prueba realmente todo nuestro trabajo de... ¿meses?

BONUS: Paso 8 Interpretación del Modelo

Puedes ver ejercicios prácticos en Python y descargar el código desde la **sección de Práctica** del Blog
ó **Instala tu ambiente de desarrollo**
Nuevo artículo: **¿Cómo hacer Webscraping? un ejemplo Python**

Ahora ya conoces los 7 pasos para crear tu propia máquina y conquistar el mundo 😊

Y si quieres crear tu propia API y desplegar en un servidor, no dejes de **leer este artículo**

Seguir Aprendiendo

Suscribe al blog y te llegará un email cada 10 ó 15 días con mi nuevo post sobre Machine Learning. Mantén tus conocimientos frescos!!

Email:

ENVIAR

Comparte el artículo:



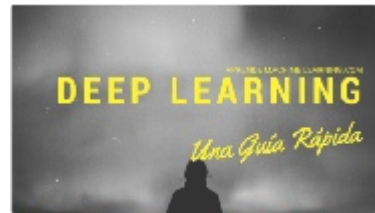
Relacionado



Análisis Exploratorio de Datos con Pandas en Python



12 Consejos útiles para aplicar Machine Learning



Aprendizaje Profundo: una Guía rápida

datos

entrenar

evaluación

Modelos

predicción

APLICACIONES DEL MACHINE LEARNING

PRINCIPALES ALGORITMOS USADOS EN MACHINE LEARNING

One comment

1.



jeremias · junio 25

Gracias!

Responder

Deja un comentario

Introduce aquí tu comentario...

Buscar

Search ...

Contacto

Suscripción

Recibe los artículos de Aprende Machine Learning en tu casilla de correo. Cada 15 días y sin Spam!

Email: Your email address here

ENVIAR

Proudly powered by WordPress | Theme: Eighties by Justin Kopepasah.

Obtén un navegador compatible para conseguir un reto reCAPTCHA.

¿Por qué tengo que hacer esto?