cómo hacer para que una empresas resuelva sus problemas usando machine learning

El mundo ha cambiado:

**Cloud** no es una solución tecnológica, es una solución de marketing.

Maquina, software, mantenimiento: lo hace la empresa de cloud y tú pagas por su uso. Así siempre dispones de la mejor máquina, etc.

**Data**: se producen más y se publican más.

**Machine** **Learning**: hay muchas librerías open source disponibles.

**IoT**

Machine learning es automatizar tareas cognitivas sencillas:

1. Generalización y relación (Supervised learning):
   * De cada elemento recopilar features o características
   * A cada elemento asignar etiquetas.
   * Buscar relación entre las características y el target

según la etiqueta que tenga habrá dos tipos.

* 1. Regression
  2. Classigication, según la

1. Comparison (Unsupervised learning):
   * Extraer características de los elementos.
   * Definir distancia entre elementos a partir de las características
   * Con esa distancia creo grupos/segmentos

2.1 Clustering: agrupar elementos parecidos.

2.2 Dimensionality reduction: reducir número de caract de los elementos

1. Reinorcement (Reinforcement learning):

EL 99% de los problemas de los negocios actualmente se resuelven con Generalización o Comparación.

Supervised learning, le llaman supervisado por que tiene etiquetas que tienen que ser asignada manualmente.

AI es un campo mucho más amplio que el machine learning.

Deep Learning es un tipo de modelo de machine learning para encontrar relaciones

Statistics: encuentra relaciones de manera matemática. Machine Learning utiliza estadística pero también fuerza bruta.

**EXAMPLE: Predecir precio de una casa**

Coger elementos del pasado y etiquetar. El target o etiqueta sería el precio de la casa.

Qué características de la casa escojo? Ese es el curro de verdad, descubrir qué características ayudarán a definir el precio de la casa.

Un modelo es la manera de encontrar la relación entre las características y el target.

Un modelo es el Arbol de decisión: coge una carac, si es más de tal ve a esta rama, si no ve a esta otra y así recursivamente. Hay que encontrar los mejores cortes, o ramas para ver la relación de las características con el precio.

Hay muchos modelos más:

* Linear regression: este es un modelo matemático pero los demás de abajo no! son modelos computacionales
* Decision tree
* Random forest
* K-neightbours
* SVM: Maquina de vector de soportes
* Neural Networks, Deep learnings

Lo modelos matemáticos abstraen los millones de datos en formulas sencillas. Los modelos computacionales guardan esos millones de datos en realidad.

Como hay muchos modelos puedes probar todos y ver cuál te da más grados de acierto.

PARAMETROS DEL MODELO

Como encuentro los mejores parámetros: ENTRENAMIENTO. Es un problema de búsqueda porque estás buscando los mejores parámetros.

Parámetros NO son las características elegidas. El modelo normalmente elige los mejores parámetros por ti.

COMO EVALUAS UN MODELO

Con métricas. Cojo un casa cuyo precio es conocido y lo meto en el modelo a ver cómo predice su precio. Usas varios modelo y ves cuál lo predice mejor. Uso métricas para medir la diferencia de valores que me han dado un modelo u otro. METRICAS son:

* Bias: media de los errores.
* MAE: valor absoluto de los errores.
* RMSE

La métrica al final la define el negocio.

Ejemplo, si tengo 1 millón de casas, uso 900000 (training set) para entrena y 100000 para testear el modelo (test set)

Después de escoger el modelo conm la mejor métrica ya lo puedo usar para predecir el precio de la casa.

Machine learning is about features, models and metrics.

CLASSIFICATION

Problem: predecir diabetes en pacientes. Es un ejemplo en que el resultado es o 1 o 0, o tienes diabetes o no. La herramientas utilizadas son las mismas que antes.

SEGMENTACION:

Problem: detectar grupos de clientes similares. No tengo etiquetas, tengo características. La segmentación está basada en buscar grupos,

Metrica: índice de silueta, dice cuán lejos está un cluster de otro.

EJEMPLO: PANADERO:

Problema, vender todas las barras en un día para que no le sobre.

Qué queremos predecir? numero de clientes. es un número

Qué tipo de problema tengo? numérico, escojo regresión.

TRABAJADORA NAVE

Objetivo: que las máquina nunca se estropeen

Pregunta a predecir: sí o no se estropeará una maquina mañana? es un problema de clasificación

MARKETING:

Predecir conversión de clientes-🡪 clasificación

MEDICO RADIOLOGO:

Predecir sí o no hay enfermedad en la radiografía 🡪 es un problema de clasificación.

Tengo un problema de business y tengo que descomponerlo para intentar resolverlo.

* La persona de business me dirá cual es el valor y la medida de éxito.
* Data sources, frecuency (cada cuanto se actualizan), legal(son públicos o privados), history
* Machine Learning task + metric
* Creo mi tabla y modelo a probar con la tabla.
* Integrar el modelo encontrado en la parte de negocio. Cuándo (a menudo) se va a usar el modelo?

CANVAS DE MACHINE LEARNING contiene los puntos de arriba, está en la presentación de diapositivas.

QUE PASA SI EL MEJOR DE TUS MODELOS NO ES SATISFACTORIO?

1. Meter más datos, líneas a tu tabla

2. Meter nuevas características.

3. Quitar características: porque puede haber características que empeoren los resultados.

Una cosa que se hace para ver qué carac no son importantes es intercambiar el contenido de dos columnas y ejecutar el modelo, si el resultado no cambia mucho es que esas caract no son importantes.

1. Probar nuevos modelos.

NOTA: Seguros y Banca no permiten por ley usar modelos difíciles de explicar, no pueden usar por ejemplo DEEP LEARNING.

Es necesario saber interpretar lo modelos

DELIVERABLES:

El data scientist entrega un modelo. O a veces simp`lemente dice lo mejor para este problema es usar este u otro modelo.

HANDS ON EXERCISES:

Weather Forecast