

Taller 19/20

ceedcv
CENTRE ESPECÍFIC
D'EDUCACIÓ A DISTÀNCIA DE
LA COMUNITAT VALENCIANA

Introducción al aprendizaje automático (Machine Learning)



¿Que es el aprendizaje automático? (1)

- Técnicas para que las computadoras
 - Aprender automáticamente
 - Sin haber sido programadas explícitamente.
 - Realizar predicciones
-

¿Que es el aprendizaje automático? (2)

- ¿Qué hacen las técnicas de aprendizaje automático?
 - **Aprenden**
 - Detectan patrones en los datos.
 - **Predicen**
 - Ajustan las acciones del programa a patrones detectados.
-

¿Que vamos a aprender?

- Aprenderemos
 - Aprendizaje supervisado.
 - Aprendizaje no supervisado.
 - Usaremos Python 3 + Scikit-learn
-

Preparando el entorno

□ **Python 3**

■ Linux Ubuntu

- `sudo apt-get install python3 python3-pip`

■ Windows

- <https://www.python.org/downloads/windows/>
- Alternativa - Anaconda
 - <https://www.anaconda.com/distribution/>

□ **Scikit-learn**

- `pip3 install -U scikit-learn`
 - Alternativa – Anaconda
 - `conda install scikit-learn`
-

Conceptos previos (1)

- **Ejemplo**

- Bla bla bla

- **Corpus:**

- Conjunto de ejemplos a aprender.
-

Conceptos previos (2)

□ **Entrenamiento:**

- Cada técnica entrena de una forma
 - Se entrena con ejemplos (aprende)
 - Cuando se entrena con ejemplos se genera un modelo.
-

Conceptos previos (3)

□ **Modelo (1)**

- Estructura que representa el conocimiento aprendido (resultado del entrenamiento)
 - Por ejemplo una red neuronal.
 - Depende de la técnica usada.
-

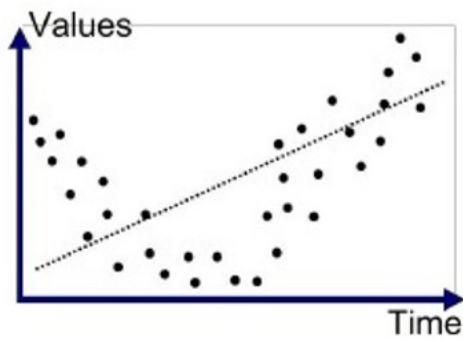
Conceptos previos (4)

□ **Modelo (2)**

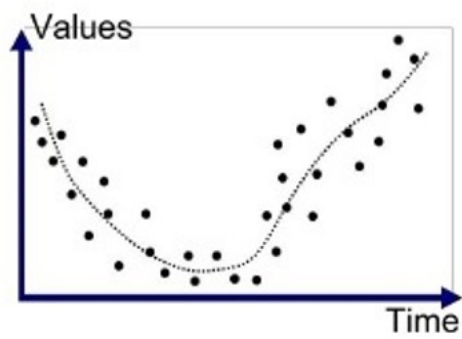
- El modelo se utiliza para predecir
 - Obtener un valor
 - Predecir precio de un piso
 - Clasificar un ejemplo
 - Indicar según edad/peso/altura si es hombre o mujer.
-

Conceptos previos (5)

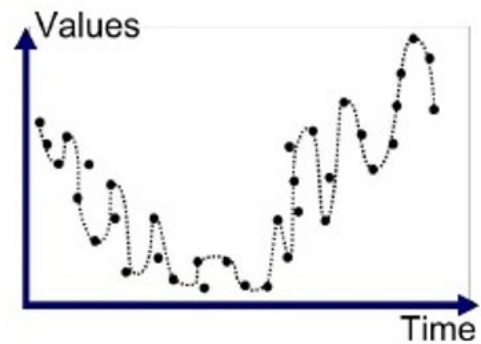
□ Entrenamiento (1)



Underfitted



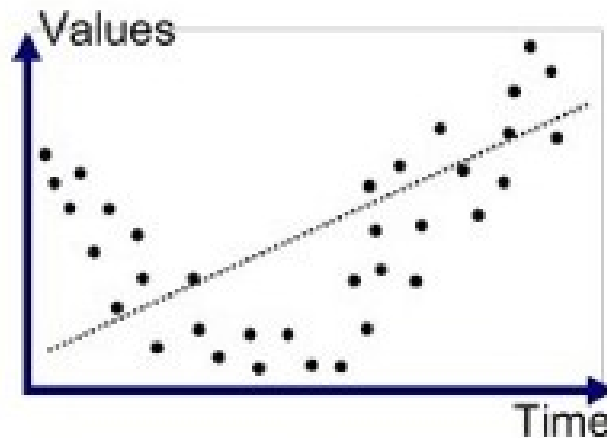
Good Fit/Robust



Overfitted

Conceptos previos (6)

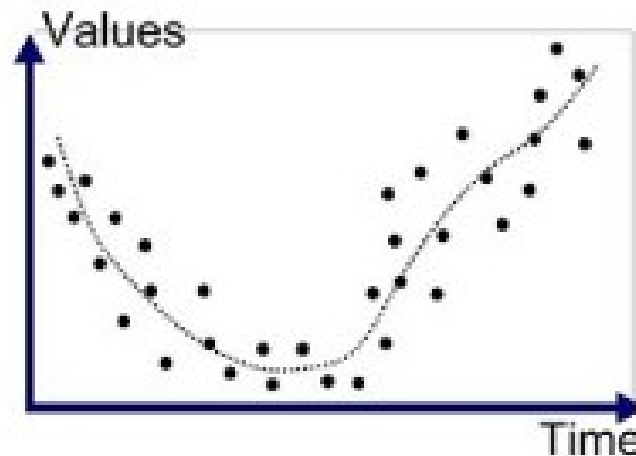
- Entrenamiento: Poco entrenado (2)



Underfitted

Conceptos previos (7)

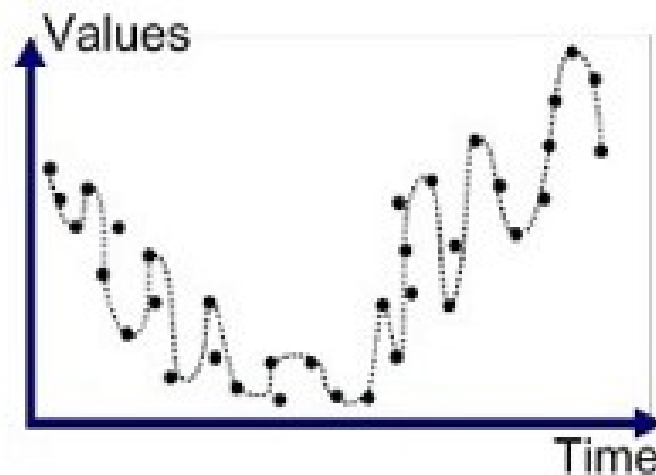
□ Entrenamiento: Adecuado (3)



Good Fit/Robust

Conceptos previos (8)

□ Entrenamiento: SobreEntrenado (4)



Overfitted

Aprendizaje supervisado (1)

- ¿Que es el aprendizaje supervisado?
 - Aprender a partir de ejemplos de los que conocemos su valor o su clasificación
-

Aprendizaje supervisado (2)

- ¿Como son los datos de entrenamiento?

- Vector de características

- | | Metros Cuadrados | Habitaciones | Baños | Precio |
|--------|------------------|--------------|-------|--------|
| Piso 1 | 100 | 3 | 2 | 100000 |
| Piso 2 | 88 | 2 | 1 | 85000 |

- Etiqueta o resultado esperado

	Metros Cuadrados	Habitaciones	Baños	Precio
Piso 1	100	3	2	100000
Piso 2	88	2	1	85000

Aprendizaje supervisado (3)

- Problemas de regresión
 - A partir de una muestra se predice el valor de una variable.

	Metros Cuadrados	Habitaciones	Baños	Precio
Piso 1	100	3	2	100000
Piso 2	88	2	1	85000
Piso 3	88	2	1	????

Aprendizaje supervisado (4)

- Problemas de clasificación
 - Se predice la clase de la muestra
 - Número de clases finito

	Peso	Altura	Sexo
Persona A	80	185	Hombre
Persona B	50	170	Mujer
Persona C	88	177	???

Aprendizaje supervisado (5)

□ **Ejemplo 1 Regresión:**

- Número de coches a una hora del día
 - Datos → Predicción
 - 8:00 → 1000
 - 09:00 → 1500
 - 10:00 → 700
 - 14:00 → 170
-

Aprendizaje supervisado (6)

□ **Ejemplo 2 Regresión:**

- Horas estudiadas, nota obtenida
 - Datos → Predicción
 - 30 horas → 6
 - 50 horas → 10
 - 35 horas → 4
 - 20 horas → 5
 - 38 horas → 7.5
-

Aprendizaje supervisado (7)

- **Ejemplo 1 Clasificación:**
 - Temperatura
 - Posibles clases (Frio/Calor)
 - $25^{\circ} \rightarrow \text{Calor}$
 - $22^{\circ} \rightarrow \text{Calor}$
 - $10^{\circ} \rightarrow \text{Frio}$
 - $0^{\circ} \rightarrow \text{Frio}$
-

Aprendizaje supervisado (8)

- **Ejemplo 2 Clasificación 2:**
 - Velocidad respecto a velocidad máxima de la vía
 - Posibles clases (Lento/Adecuado/Peligro)
 - 50, 120 → Lento
 - 40, 50 → Adecuado
 - 80, 60 → Peligro
 - 100, 120 → Adecuado
 - 140, 120 → Peligro
-

Caso Estudio Regresión

- Caso de estudio Regresión
 - Algoritmo: Regresión lineal mínimos cuadrados
 - Corpus simple precio pisos en relación a metros cuadrados.
 - Ver "***RegresionPisos.py***"
 - Saber mas

<http://www.aprendemachinelearning.com/regresion-lineal-en-espanol-con-python/#more-5722>

Caso Estudio Clasificación (1)

- Caso de estudio Clasificación (1)
 - Juego que juega aleatoriamente a 3 en raya.
 - Haremos que la máquina aprenda de distintas formas.
-

Caso Estudio Clasificación (2)

- **Paso 1: Generar corpus (1)**
 - Programa que juega millones de partidas automáticamente
 - Ver "***GeneraCorpus3EnRaya.py***"
-

Caso Estudio Clasificación (3)

- **Paso 1: Generar corpus (2)**
 - De cada partida, guardamos para todos los movimientos, como estaba el tablero y que movimiento se hizo.
-

Caso Estudio Clasificación (4)

□ **Paso 1: Generar corpus (3)**

- **Datos:** Estado del tablero y que movimiento se hizo.
 - **Clasificación de los datos**
 - Si el que hizo el movimiento al finalizar la partida acabó ganando o no
-

Caso Estudio Clasificación (5)

□ Paso 2: Entrenamiento (1)

- Entrenamos con el corpus generado y guardamos entrenamiento para ser usado en otros ficheros
 - Árbol de decisión
 - Ver fichero "***EntrenaModeloTree.py***"
 - *SVM (Maquinas de vectores de soporte)*
 - Ver fichero "***EntrenaModeloSVM.py***"
-

Caso Estudio Clasificación (6)

□ Paso 2: Entrenamiento (2)

- Uso de puntuación de la calidad de la predicción
 - Algoritmo usado: Árbol de decisión
 - Uso de clasificador y seleccionando aleatorio entre candidatos
 - Algoritmo usado: SVM
-

Caso Estudio Clasificación (7)

□ Paso 3: Comparativa

- Usamos el entrenamiento para jugar y probar efectividad
 - Con score ***"3EnRayaJuego.py"***
 - Sin score
"3EnRayaJuegoSinScore.py"

Caso Estudio Clasificación (8)

- Caso de estudio Clasificación (4)
 - Saber más
 - <http://www.aprendemachinelearning.com/principales-algoritmos-usados-en-machine-learning/>
-

Aprendizaje no supervisado (1)

- ¿Que es el aprendizaje no supervisado?
 - No hay entrenamiento previo.
 - Busca agrupaciones
 - Eso no implica que dichas agrupaciones tengan sentido o utilidad
 - El modelo se va ajustando según las observaciones.
-

Aprendizaje no supervisado (2)

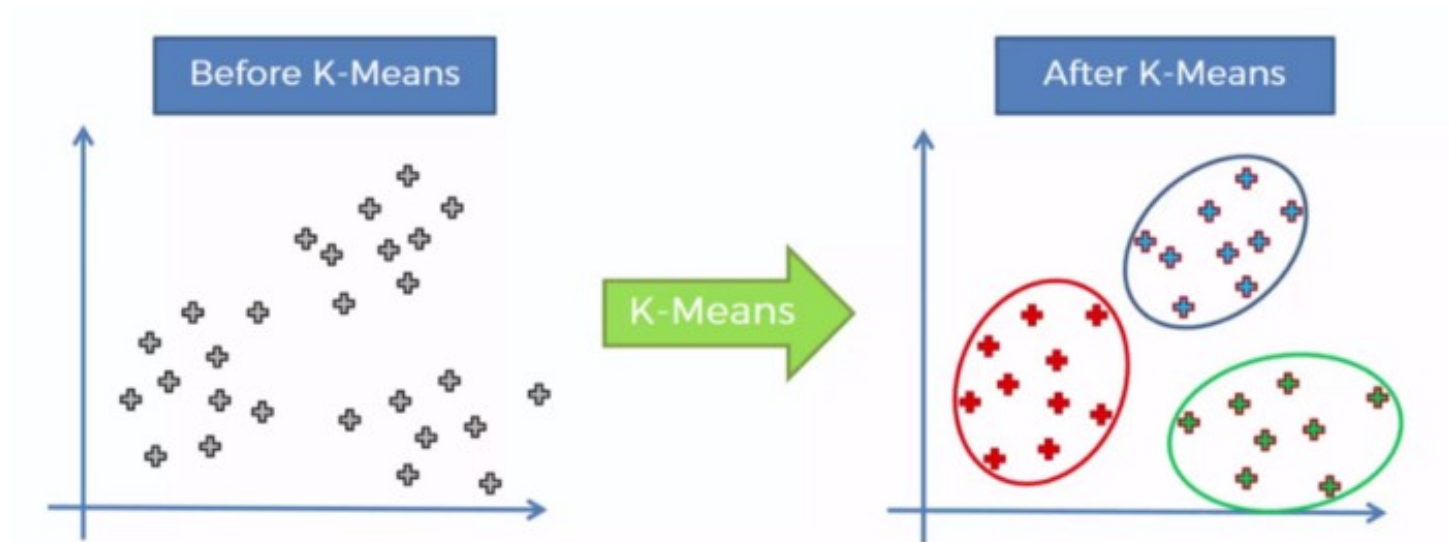
- Ejemplo: Algoritmo K-means (1)
 - <https://es.wikipedia.org/wiki/K-medias>
 - Elegimos cuantos K grupos queremos
 - Se eligen K ejemplos aleatorios y se establecen como centros.
 - Procedemos uno a uno a ir añadiendo el resto de ejemplos de entrenamiento
-

Aprendizaje no supervisado (3)

- Ejemplo: Algoritmo K-means (2)
 - Cada nuevo ejemplo introducido
 - Se clasifica asocia a su centro más cercano
 - Algoritmo re-calcula la posición de los centros
 - Se vuelven a asignar todos los ejemplos a su centro más cercano
-

Aprendizaje no supervisado (3)

□ Ejemplo: Algoritmo K-means (3)



- <http://exponentis.es/ejemplo-de-clustering-con-k-means-en-python>

Aprendizaje no supervisado (3)

□ Caso de estudio (1)

- Datos: número de goles y de asistencias en una temporada

	Goles	Asistencias
C. Ronaldo	24	5
Messi	25	13
Xavi	3	16
Pique	4	2

- Fichero "***NoSupervisadoKMeans.py***"
-

Aprendizaje no supervisado (4)

□ **Caso de estudio (2)**

- Ajustamos el modelo en base a observaciones
 - Elegimos cuantos grupos distintos
 - Observamos datos agrupados y les buscamos sentido
-

Aprendizaje no supervisado (5)

□ **Caso de estudio (3)**

■ **Observaciones y posibles grupos**

- Grupos solo goleadores
 - Grupos solo asistentes
 - Grupos goleadores y asistentes
 - Grupos ni goleadores ni asistentes
-

Para saber más

□ Enlaces interesantes

- Mi curso favorito!!
- <https://es.coursera.org/learn/machine-learning>
- Otros enlaces interesantes
- <https://medium.com/datos-y-ciencia/introduccion-al-machine-learning-una-gu%C3%ADa-desde-cero-b696a2ead359>
- <https://relopezbriega.github.io/blog/2015/10/10/machine-learning-con-python/>
- <https://eu.udacity.com/course/intro-to-artificial-intelligence--cs271>
- <https://www.coursera.org/lecture/machine-learning-with-python/introduction-to-machine-learning-zSm8k>