

Introducción al Aprendizaje de Máquina

Mauricio A. Álvarez

Modelos probabilísticos profundos
AIR Institute

Contenido

Introducción

Definiciones

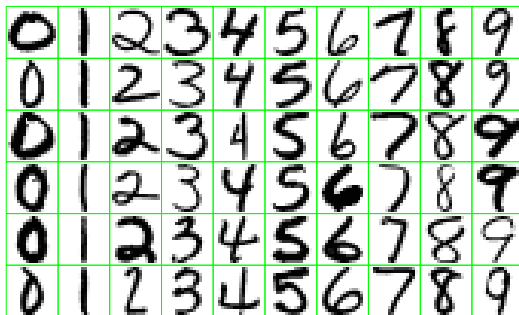
Análisis de Patrones

Aprendizaje de Máquina ó Aprendizaje Estadístico

- ❑ Este curso es sobre *aprendizaje a partir de datos*.
- ❑ Se estudian diferentes algoritmos que permiten resolver diferentes problemas de predicción.
- ❑ El algoritmo se diseña con base en un modelo matemático o función, y una base de datos.
- ❑ Se desea extraer conocimiento a partir de datos.

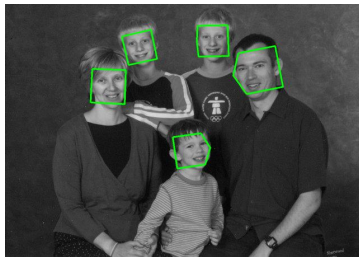
Ejemplos de problemas de predicción

Reconocimiento de dígitos manuscritos



Ejemplos de problemas de predicción

Detección y reconocimiento de rostros.



Ejemplos de problemas de predicción

Predecir la edad de una persona que mira un video particular en YouTube.



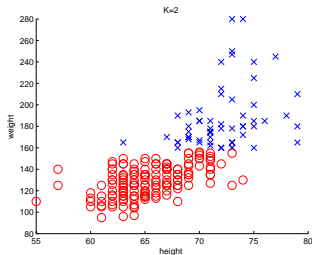
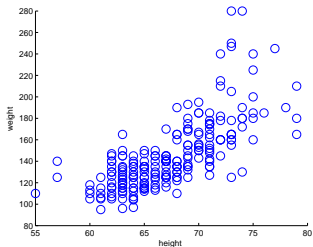
Ejemplos de problemas de predicción

Mercado de acciones



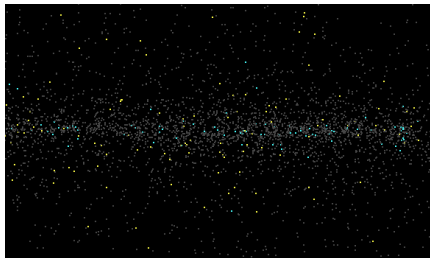
Ejemplos de problemas de predicción

Agrupamiento



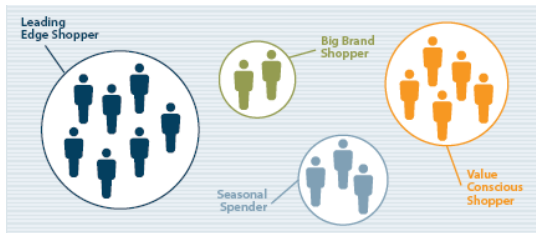
Ejemplos de problemas de predicción

Autoclass



Ejemplos de problemas de predicción

Segmentación de clientes en comercio electrónico



Ejemplos de problemas de predicción

Sistemas de recomendación


Customers Who Bought This Item Also Bought

Page 1 of 17






Machine Learning: A Probabilistic...
› Kevin P. Murphy
★★★★★ 35
Hardcover
\$81.71 ✓Prime



The Elements of...
Trevor Hastie
★★★★★ 40
#1 Best Seller in
Bioinformatics
Hardcover
\$84.04 ✓Prime



Probabilistic Graphical Models: Principles and...
› Daphne Koller
★★★★★ 26
Hardcover
\$99.75 ✓Prime



Machine Learning with R
Brett Lantz
★★★★★ 26
Paperback
\$49.49 ✓Prime



An Introduction to...
› Gareth James
★★★★★ 37
#1 Best Seller in
Mathematical & Statistical...
Hardcover
\$75.99 ✓Prime



Reinforcement Learning: An Introduction...
› Richard S. Sutton
★★★★★ 17
Hardcover
\$64.60 ✓Prime





Ejemplos de problemas de predicción

Image inpainting



Ejemplos de problemas de predicción

Google's Stunning Win For...


www.forbes.com/sites/parmyolson/2016/03/12/google-wins-a-i-alpha-go-beats-human-champion-in-third-match/#3f3bfcb4722

Forbes / Tech


THE LITTLE BLACK BOOK OF BILLIONAIRE SECRETS

MAR 12, 2016 @ 09:47 AM 2,920 VIEWS

Google's Stunning Win For A.I.: AlphaGo Beats Human Champion In Third, Crucial Match



Parmy Olson
FORBES STAFF
I cover agencies and
innovators in tech.
FOLLOWING ON SCHEMATA.COM
f t s
FULL BIO >




Journalists watch a big screen displaying live footage of the third game of the Google DeepMind Challenge Match between Lee Sedol, one of the greatest modern players of the ancient board game Go, and the Google-developed supercomputer AlphaGo at a hotel in Seoul on March 12, 2016. A Google-developed computer programme took an unassailable 3-0 lead in its match-ups with a South Korean Go grandmaster on March 12 — marking a major breakthrough for a new style of "intuitive" artificial intelligence. (Photo: JUNG YEON-JU/AP/ Getty Images)

Google **GOOGL** +5.97% 's AlphaGo computer program has won its third, pivotal match against a champion player of Go, the ultra-complex strategic game that till now was seen as one of the last vestiges of human superiority over computers in games.

AlphaGo has now won the overall best-out-of-five series of games in three straight sets, and in all three games it displayed extraordinary feats of intuition and creativity which stunned observers.

"AlphaGo was stronger than anyone had expected," said Lee Hyunwook, a professional Go player who acted



Microsoft Cloud

See how the Digital Crimes Unit helps protect the cloud.

Get free demo

Contenido

Introducción

Definiciones

Análisis de Patrones

Definiciones básicas

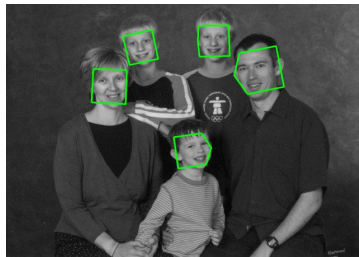
- ❑ **Conjunto de entrenamiento** ($\{\mathbf{X}, \mathbf{t}\}$): conjunto de N dígitos $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ con sus correspondientes etiquetas $\{t_1, \dots, t_N\}$, empleados para sintonizar los parámetros de un modelo de predicción.
- ❑ **Entrenamiento**: el resultado de ejecutar el algoritmo de aprendizaje de máquina puede ser expresado como una función $y(\mathbf{x}, \theta)$ que toma un dígito \mathbf{x} y entrega una salida codificada de la misma manera que \mathbf{t} .
- ❑ **Validación**: probar la función obtenida con un conjunto diferente de dígitos a los utilizados en el entrenamiento (conjunto de validación).
- ❑ **Generalización**: habilidad de clasificar correctamente dígitos de validación.

Aprendizaje supervisado y no supervisado

- **Aprendizaje supervisado:** se conocen los vectores de características (\mathbf{X}), y sus correspondientes etiquetas de salida (\mathbf{t}).
 - La variable de salida es discreta: *clasificación*.
 - La variable de salida es continua: *regresión*.
- **Aprendizaje no supervisado:** sólo se conocen los vectores de características (\mathbf{X})
 - Descubrir grupos de datos similares: *agrupamiento*.
 - Determinar la distribución de los datos: *estimación de densidad*.
 - Proyectar los datos a un espacio de menor dimensión: *reducción de dimensionalidad*.
- **Otro tipo de aprendizaje:** semi-supervisado, aprendizaje activo, aprendizaje de múltiples etiquetas, aprendizaje por refuerzo.

Aprendizaje supervisado: aplicaciones (I)

- ❑ Reconocimiento de dígitos manuscritos.
- ❑ Detección y reconocimiento de rostros.



Aprendizaje supervisado: aplicaciones (II)

- ❑ Las anteriores son aplicaciones de clasificación.
- ❑ Aplicaciones de regresión incluyen predecir la edad de una persona que mira determinado video en YouTube.



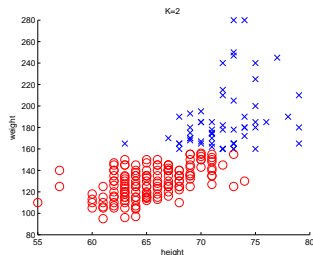
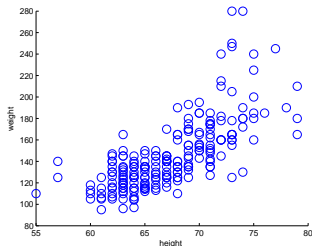
Aprendizaje supervisado: aplicaciones (III)

- Otra aplicación de regresión es predecir el precio futuro de una acción en el mercado de valores, dadas las condiciones actuales del mercado y otra posible información adicional.



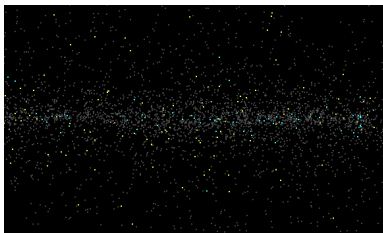
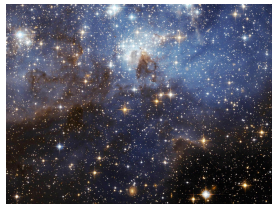
Aprendizaje no supervisado: aplicaciones (I)

□ Agrupamiento



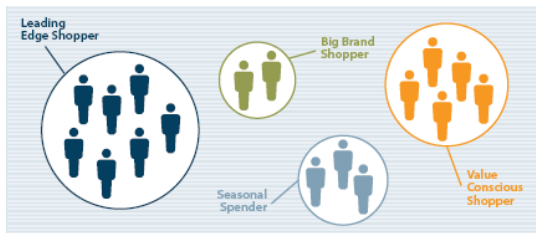
Aprendizaje no supervisado: aplicaciones (II)

- En astronomía, el sistema **autoclass** descubrió un nuevo tipo de estrella basándose en el agrupamiento de mediciones astrofísicas.



Aprendizaje no supervisado: aplicaciones (III)

- En comercio electrónico (e-commerce) es común agrupar usuarios en grupos, basándose en su compras y las páginas que navega, y luego enviando publicidad personalizada a cada grupo.

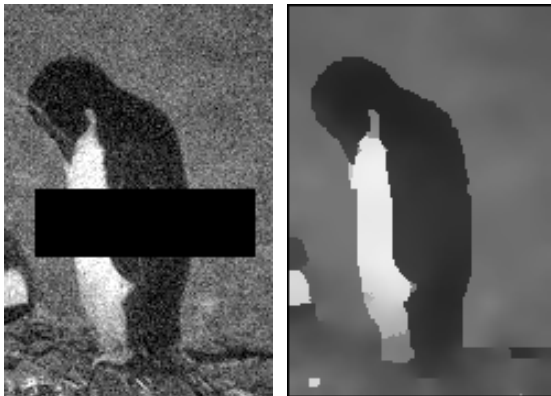


Aprendizaje no supervisado: aplicaciones (IV)

- ❑ Llenado de matrices (matrix completion).
- ❑ Algunas veces se tienen datos faltantes, es decir, variables cuyos valores no se conocen.
- ❑ Por ejemplo, se realizó una encuesta y algunos de los encuestados no contestaron ciertas preguntas.
- ❑ Se puede tener una red de sensores, y algunos de ellos pueden fallar.
- ❑ Se debe realizar una imputación de esos valores.

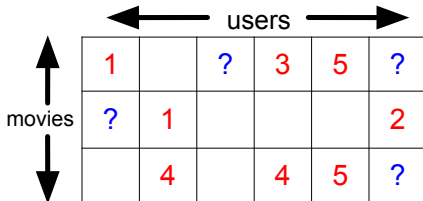
Aprendizaje no supervisado: aplicaciones (V)

- Image inpainting: algunas imágenes pueden tener oclusiones.



Aprendizaje no supervisado: aplicaciones (VI)

- ❑ Collaborative filtering.
- ❑ Predecir cuáles películas una persona va a querer ver basándose en cómo esa persona y otras personas diferentes han calificado películas que ellos ya hayan visto.
- ❑ Netflix: matriz de 18k películas y 500k usuarios, aprox. 10^9 entradas, pero sólo se observa 1%.



	1		?	3	5	?
?	1					2
	4		4	5		?

Aprendizaje no supervisado: aplicaciones (VII)

Sistemas de recomendación.



Customers Who Bought This Item Also Bought

Page 1 of 17

Machine Learning: A Probabilistic...
› Kevin P. Murphy
★★★★★ 35
Hardcover
\$81.71 ✓Prime

The Elements of...
Trevor Hastie
★★★★★ 40
#1 Best Seller in Bioinformatics
Hardcover
\$84.04 ✓Prime

Probabilistic Graphical Models: Principles and...
› Daphne Koller
★★★★★ 26
Hardcover
\$99.75 ✓Prime

Machine Learning with R
Brett Lantz
★★★★★ 26
Paperback
\$49.49 ✓Prime

An Introduction to...
› Gareth James
★★★★★ 37
#1 Best Seller in Mathematical & Statistical...
Hardcover
\$75.99 ✓Prime

Reinforcement Learning: An Introduction...
› Richard S. Sutton
★★★★★ 17
Hardcover
\$64.60 ✓Prime

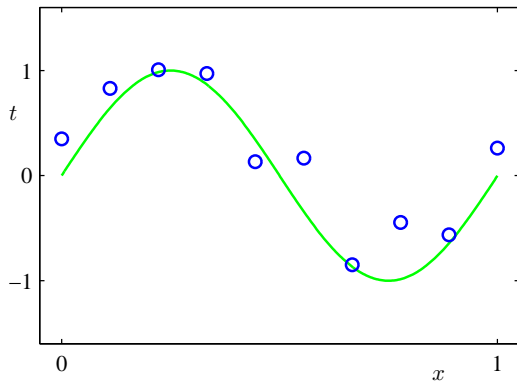
Contenido

Introducción

Definiciones

Análisis de Patrones

Ejemplo Regresión (I)



- ❑ **Regresión:** supongamos una función conocida $\sin(2\pi x)$ con ruido aleatorio incluido en la variable objetivo \mathbf{t} .
- ❑ Conjunto de entrenamiento: $\mathbf{x} \equiv \{x_1, \dots, x_N\}^\top$, $\mathbf{t} \equiv \{t_1, \dots, t_N\}^\top$.

Ejemplo Regresión (II)

- ❑ Objetivo: usar el conjunto de entrenamiento para hacer predicciones \hat{t} para algún valor nuevo de \hat{x} .
- ❑ **Dificultad:** generalizar $\sin(2\pi x)$ a partir de un conjunto finito de datos.
- ❑ Debido al ruido hay incertidumbre acerca del verdadero valor de \hat{t} .

Función polinomial (I)

- Usando como modelo de predicción una función polinomial,

$$y(x, \mathbf{w}) = w_0 + w_1 x + \dots + w_M x^M = \sum_{j=0}^M w_j x^j,$$

donde $\mathbf{w} \equiv \{w_0, w_1, \dots, w_M\}$.

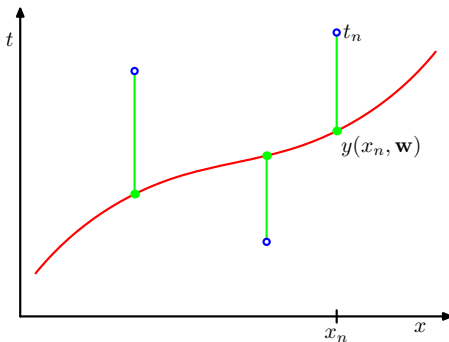
- Nótese que la función es lineal respecto a \mathbf{w} .
- El proceso de entrenamiento consiste en encontrar los coeficientes \mathbf{w} , que mejor ajusten el polinomio al conjunto de entrenamiento.

Función polinomial (II)

- Esto se realiza minimizando una función de error

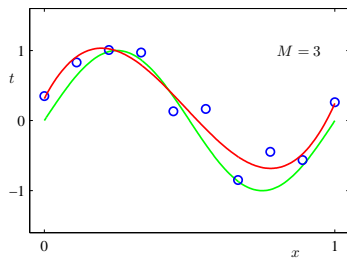
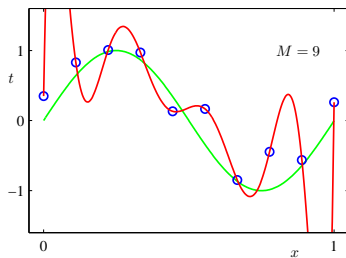
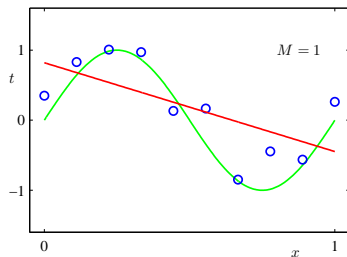
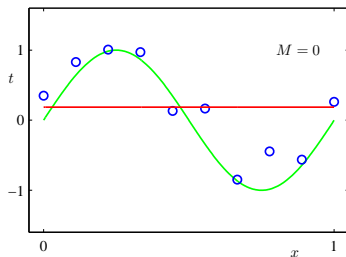
$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{y(x_n, \mathbf{w}) - t_n\}^2.$$

- La minimización del error tiene solución única \mathbf{w}^* . El polinomio resultante estará dado por $y(x, \mathbf{w}^*)$.



Selección del modelo

Una pregunta natural: cómo escoger el orden del polinomio, M ?

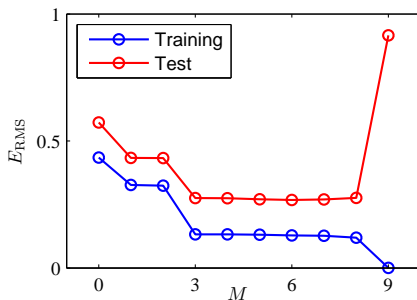


Validación

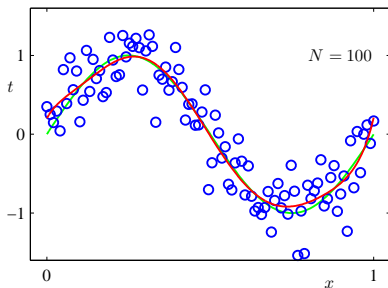
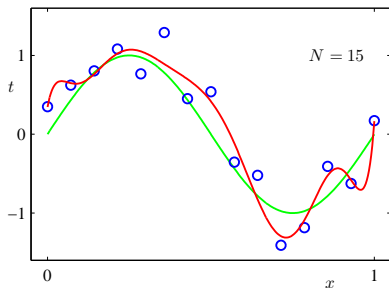
- Para lograr una buena generalización, verificar $E(\mathbf{w}^*)$ sobre un conjunto de validación usando

$$E_{\text{RMS}} = \sqrt{2E(\mathbf{w}^*)/N}.$$

- El error RMS permite comparar errores para conjuntos de diferentes tamaños.

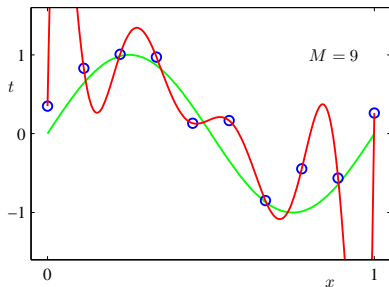


Modelo en función de N ($M = 9$).



Entre más grande el conjunto de entrenamiento, más complejo el modelo que se puede usar.

Regularización (I).



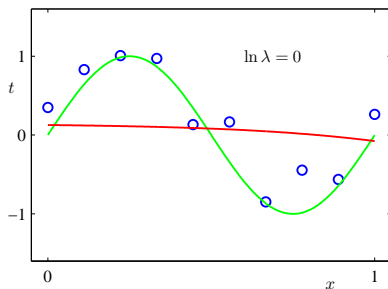
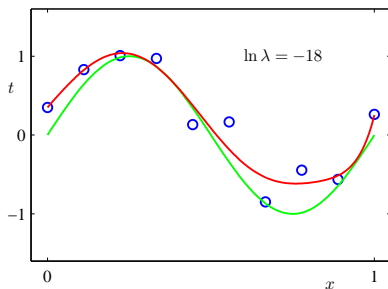
	$\ln \lambda = -\infty$	$\ln \lambda = -18$	$\ln \lambda = 0$
w_0^*	0.35	0.35	0.13
w_1^*	232.37	4.74	-0.05
w_2^*	-5321.83	-0.77	-0.06
w_3^*	48568.31	-31.97	-0.05
w_4^*	-231639.30	-3.89	-0.03
w_5^*	640042.26	55.28	-0.02
w_6^*	-1061800.52	41.32	-0.01
w_7^*	1042400.18	-45.95	-0.00
w_8^*	-557682.99	-91.53	0.00
w_9^*	125201.43	72.68	0.01

Se puede regularizar la función de error para prevenir que \mathbf{w} tome valores grandes

$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{y(x_n, \mathbf{w}) - t_n\}^2 + \frac{\lambda}{2} \|\mathbf{w}\|^2,$$

donde λ es el término de regularización.

Regularización (II).



	$\ln \lambda = -\infty$	$\ln \lambda = -18$	$\ln \lambda = 0$
w_0^*	0.35	0.35	0.13
w_1^*	232.37	4.74	-0.05
w_2^*	-5321.83	-0.77	-0.06
w_3^*	48568.31	-31.97	-0.05
w_4^*	-231639.30	-3.89	-0.03
w_5^*	640042.26	55.28	-0.02
w_6^*	-1061800.52	41.32	-0.01
w_7^*	1042400.18	-45.95	-0.00
w_8^*	-557682.99	-91.53	0.00
w_9^*	125201.43	72.68	0.01