機器學習 - NTUDAC James Yeh



James Yeh

- Yahoo 軟體工程師(Backend, ML)
- 台大電機所
- 網站開發、資料分析、機器學習
- Python 資料分析與機器學習入門
- yehjames23@gmail.com

今天我們會



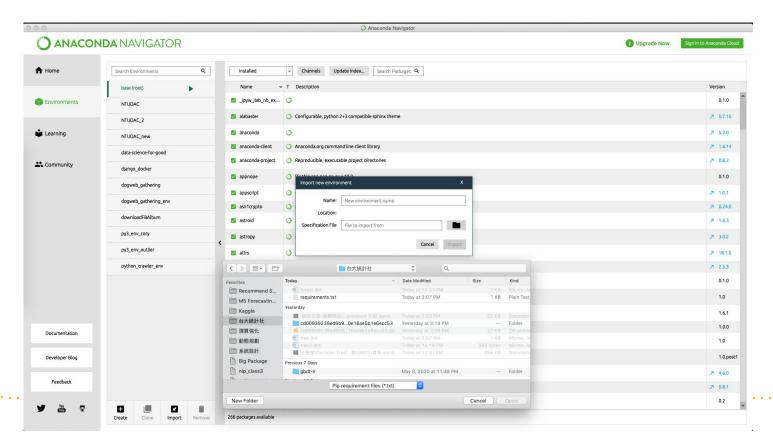
環境安裝

重裝一個乾淨的conda環境(避免套件衝突)

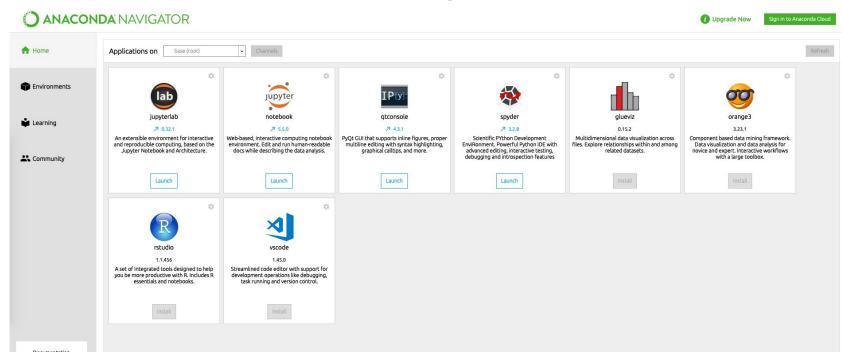
在你的終端機:

conda create --name NTUDAC python=3.6 conda activate NTUDAC pip install -r requirements.txt conda deactivate

或是從Anaconda navigator



選擇環境 開啟Jupyter



機器學習初探

機器學習可以做什麼呢?

Python 機器學習以及Scikit-learn介紹

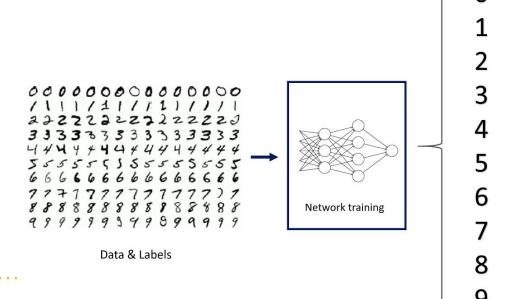
垃圾郵件分類



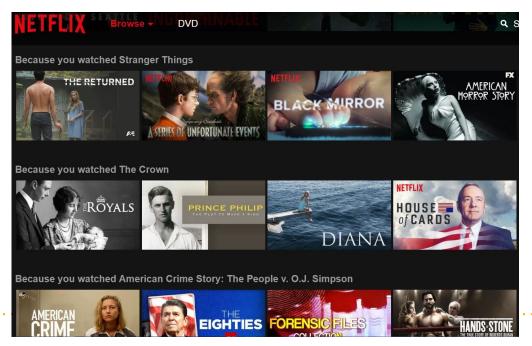
人臉辨識



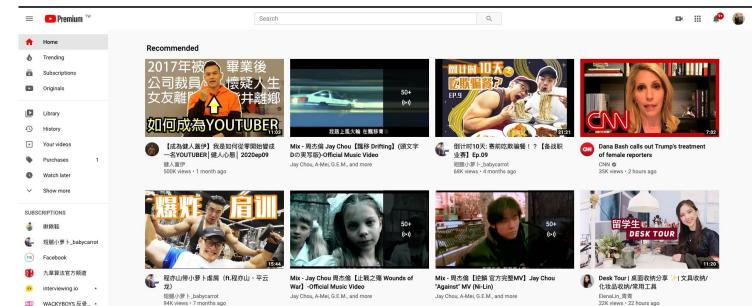
數字辨識



推薦系統(NETFLIX)



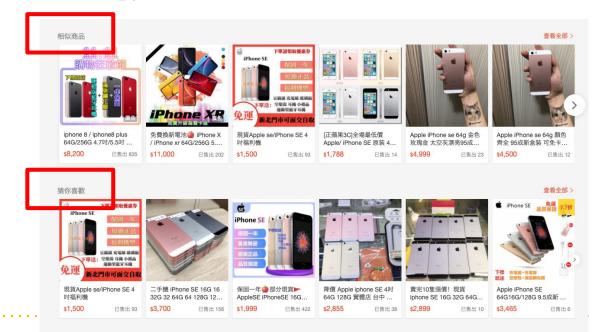
推薦系統(Youtube)



推薦系統(FB)



推薦系統(蝦皮)



真實機器學習案例(Quora)

Yoshua Bengio: Where is deep learning research headed?



Yoshua Bengio, My lab has been one of the three that started the deep learning approach, bac...

14.9k Views • Upvoted by Alberto Bietti • Xavier Amatriain • 23 others you follow

Answer featured in The Huffington Post.

Yoshua has 43 endorsements in Deep Learning

Research is by definition exploratory, which means that (a) we do not know what will work and (b) we need to explore many paths, we need a lot of diversity of research directions in the scientific community. So I can only tell you about my current gut feelings and visions of where I see important challenges and opportunities that appeal to my personal aesthetics and instincts. Here are some elements of this:

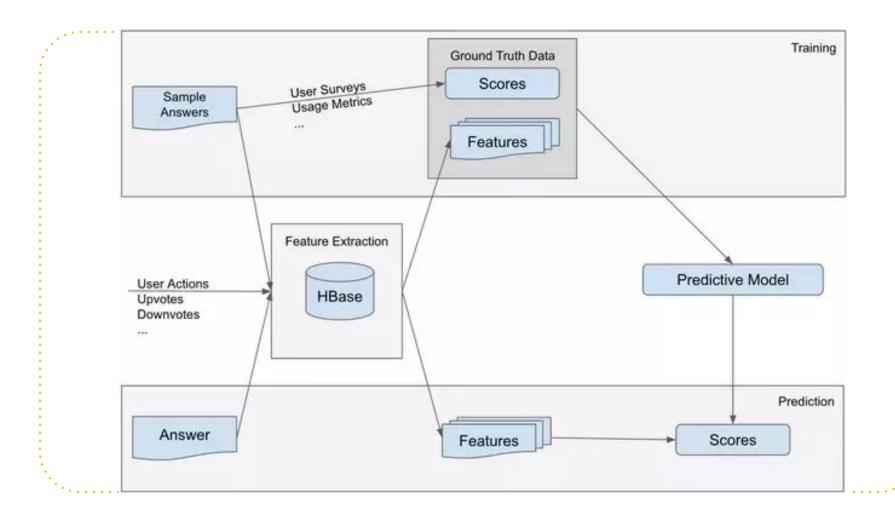
* unsupervised learning is crucial and we do not do it right yet (there are many arguments I and others have written and talked about to justify this)

* doon learning receased is likely to continue its expansion from traditional pattern

Upvote 249

Downvote Comments 2+ Share 5

...



AI發展

AI發展已成為不可逆的趨勢

- 人臉辨識
- 推薦系統
- •

各行各業開始使用 大數據+模型節省人力,物力,時間 ☑ udn 聯合新聞網 (新聞發布)

自動化與無人化的軍事革新:軍用無人載具的下一階段發展

無人載具進一步快速發展的關鍵,在於結合AI人工智慧。 ... 戰機飛行,由人工智慧系統負責絕大部份的飛行工作,讓一架傳統戰機上的飛行員 ... 由此可見,軍用無人機的發展已跨入空戰的領域中,自動化飛行將取代人工遙控的 1 day ago



UDN 聯合新聞網 (新聞發布)

未來還需人類工作嗎?迎AI時代,與機器共存的預習

《人類未來方程式》中不迴避地一再提起人工智慧將取代基本勞力的主張,但同時也強調人類與機器合作能創造更高的生產力,並帶給人類更好的生活



3 weeks ago

N 科技新報 TechNews

離全面自動化雖還早,但機器人勞力確實越來越便宜了

... 不斷發展的人工智慧所威脅甚至反噬的情節在文學作品中層出不窮,機器 人取代重複性的勞動工作也是公認的自動化高度發展後可以預見的後果。





入 民報 (新聞發布)

【民報】【專欄】人工智慧拯救建築業

建築公司開始使用AI和機器學習來更好地規劃工作中勞力和機械的分配。機器人... 儘管有大量失業預測,人工智慧不太可能取代人力資源。相反,它... 5 days ago



何時不用AI



Allie K. Miller • 3rd+
Forbes Al Innovator of the Year | Artificial Intelligence at Amazon | LinkedIn To...
1mo • 🚱

When is machine learning NOT a good idea?

- No data (due to budget or access)
- A rules-based solution works
- Low ROI for your business
- Vou just want something cool
- No tolerance for mistakes
- No one to maintain it

AI學習門檻(從理論出發不適合初學者)

5. Logistic regression is an important binary classification technique in machine learning that builds off of the concepts of linear regression. Recall that in linear regression, there is a set of predictor variables (or features) $\mathbf{a}_i \in \mathbb{R}^d$, $i=1,\ldots,n$, with corresponding outcome variables $b_i \in \mathbb{R}$, $i=1,\ldots,n$ and our aim is to fit a linear model to the data that minimizes a quadratic loss, namely

$$\min_{\mathbf{x} \in \mathbb{R}^d} \frac{1}{2} \sum_{i=1}^n \left(\langle \mathbf{a}_i, \mathbf{x} \rangle - b_i \right)^2.$$

It can be easily seen that the solution to the above least-squares is the same as

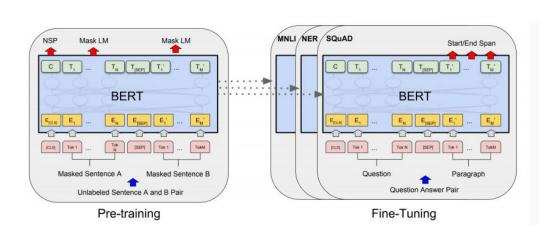
$$\min_{\mathbf{x} \in \mathbb{R}^d} f(\mathbf{x}) \triangleq \sum_{i=1}^{n} \phi(\langle \mathbf{a}_i, \mathbf{x} \rangle) - b_i \langle \mathbf{a}_i, \mathbf{x} \rangle,$$

where $\phi(t)=t^2/2$. Logistic regression can be thought of as a modification of linear regression in two ways: first, the outcome variables are binary representing the two classes, i.e., $b_i \in \{0,1\}$, $i=1,\ldots,n$, and second, the least-squares loss is replaced with a logistic loss, i.e., $\phi(t)=\ln(1+e^t)$, where "ln" is natural logarithm. Logistic regression is thus the problems of finding the parameter $\mathbf x$ that minimizes this new loss (note that unlike linear regression, there is no closed form solution for logistic regression and one has to resort to optimization algorithms). In machine learning, this phase is often referred to as "training". After the training phase, we now have a way of classifying a new input, that is not part of your training data, i.e., predicting the label of an out-of-sample data. This phase in machine learning is often called "testing" or "generalization". More specifically, using our trained logistic model and given any new data $\mathbf a$, we can obtain the probability of $\mathbf a$ belonging to either class 0 or 1 as

$$\mathbf{Pr}(y=0 \mid \mathbf{a}) = \frac{1}{1+e^{\langle \mathbf{a}, \mathbf{x} \rangle}}, \quad \text{and} \quad \mathbf{Pr}(y=1 \mid \mathbf{a}) = \frac{1}{1+e^{-\langle \mathbf{a}, \mathbf{x} \rangle}}.$$



AI學習門檻(許多抽象的概念跟前因後果)





實際解決問題or面試

- 1. 定義問題跟分析問題重要度大於ML model
- 2. **合理的代理問題**,(你怎麼量化user的滿意度?)
- 3. model**用最簡單的再慢慢加上去即可**(Baseline model)
- 4. 實際題目(第三堂課):
 - a. 如何設計Facebook news feed ranking
 - b. 如何幫Facebook設計一個檢測NSFW content的系統

AI相關職業

Machine learning/ Al researcher(PhD)

Data scientist(PhD ~ SQL Boy)

Machine learning/ AI Engineer

Business analyst

人類學習-觀察業界典範

Shopify

Shopify介紹

Govale

shopify 介紹



Shopify介紹:給台灣、香港與新加坡用戶參考的評價

https://iknowwang.com > 電商平台比較 > shopify ▼ Translate this page Shopify對於線上販賣就像是會計師對於Microsoft excel一樣。大部分的電子商務商店擁有著都知 道如何使用它。 建立網路商店真的這麼簡單?我們建議你免費試用看...

一篇文章带你全面了解Shopify开店| Shopify教程

https://www.exportb2c.com > about-shopify - Translate this page Dec 6, 2018 - Shopify是什么? Shopify是一个能让用户自主管理在线商店的平台,通过介绍 Shopify的由来及发展, Shopify开店优势, Shopify开店的费用以及...

師法亞馬遜,十張圖剖析電商隱形冠軍Shopify | 數位時代

https://www.bnext.com.tw > article > shopify-ecommer... * Translate this page 師法亞馬遜,十張圖剖析電商隱形冠軍Shopify, 2017.08.15 by, IEObserve國際經濟觀察 作者簡介. IEObserve國際經濟觀察 查看更多文章. 關心國際經濟、科技和商業 ...

5招選對開店平台! SHOPLINE 和Shopify 開店平台大比拼...

https://shopline.tw > 首頁 > 開店教學 ▼ Translate this page

Aug 2, 2018 - 今天SHOPLINE 就和Shopify來個比較,看看那個適合你作為電商創業的開始! ... 適 合自己的平台是哪種,下半部會比較全球兩大開店平台SHOPLINE 和Shopify 的十個功能。 ... 8 個 遠端工作線上工具介紹,在家工作、防疫一次搞定!

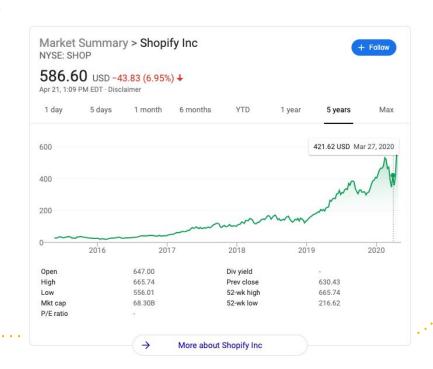
People also search for shopline評價ptt SHOPLINE POS shopify台灣金流 Wix 購物車 shopline教學 shopify教學

Shopify 中文版來了! - Shopify Taiwan

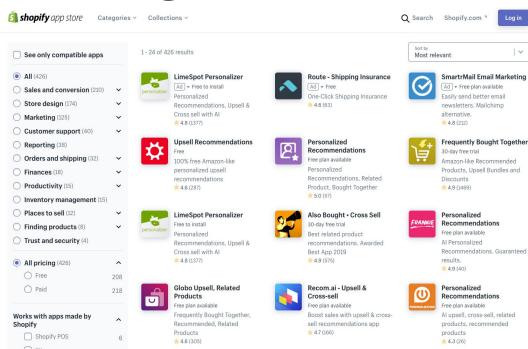
https://www.shopifytaiwan.com > shopify中文版來了! ▼ Translate this page 號稱地表最強的線上開店平台Shopify,現在官方推出Shopify 繁體中文版操作手冊,終於可以把 Google 翻譯丟一邊,讓開店過程更簡單! Shopify 說明中心引導你從 ...

20個最佳的Shopify主題和漂亮的電商界面設計

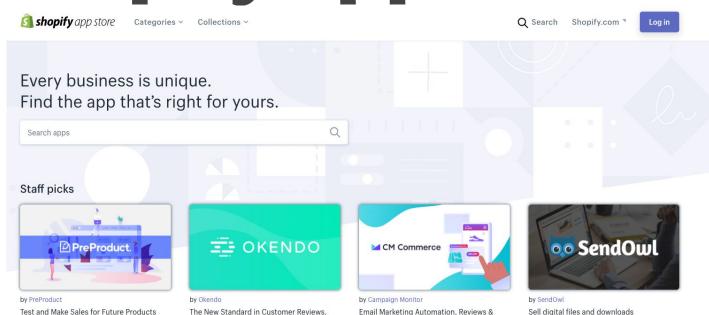
https://webdesign.tutsplus.com > zh-hant > articles > 20-... Translate this page Nov 20, 2019 - 你可以在Envato Market上我們提供的Shopify電商主題中找到所有符合... 問到的問 題、添加了更為清晰的產品介紹、添加選擇性文字到你的圖片、...



Shopify App store



Shopify App store



Popups. (Prev Conversio)

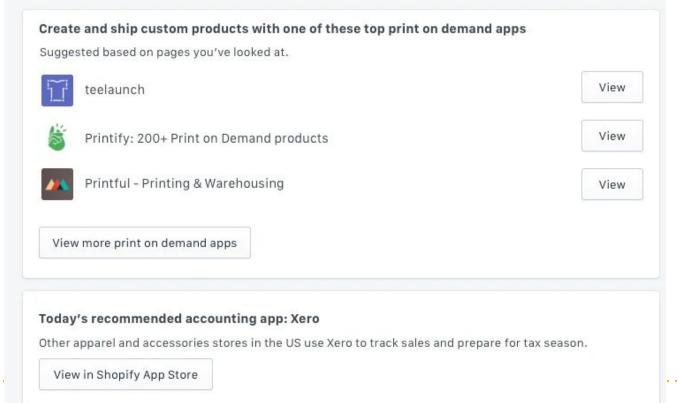
Shopify 推薦系統介紹



https://engineering.shopify.com/blogs/engineering/how-shopify-uses-recommender-systems-to-empower-entrepreneurs

including web, mobile, social media, marketplaces, brick-and-mortar locations, and pop-up shops.

Shopify 推薦系統-解釋性

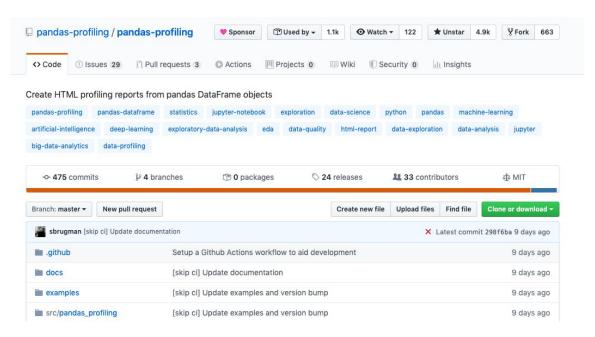


Shopify 推薦系統-結果

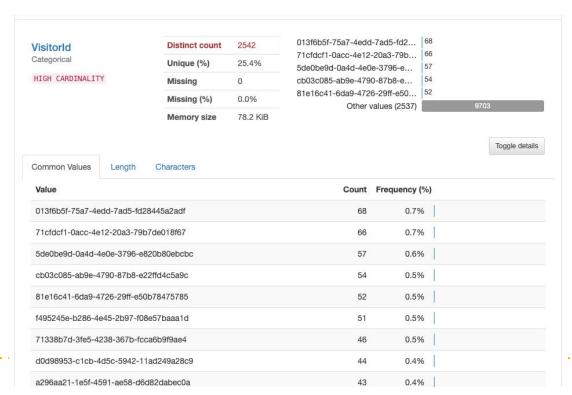
 Merchants receiving personalized app recommendations on the Shopify App Store had a 50% higher app install rate compared to those who didn't receive recommendations

Data preprocess & analysis

資料探索神器



客戶行為數



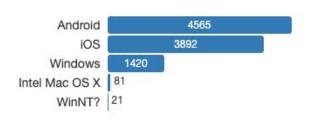
裝置行為數

是手機App步驟比較複雜還是大家喜歡用手機App?

OperationSystem

Categorical

Distinct count	6
Unique (%)	0.1%
Missing	8
Missing (%)	0.1%
Memory size	78.2 KiB



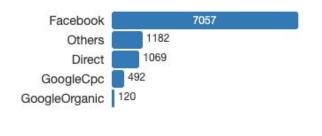
Toggle details

流量來源

TrafficSourceCateg...

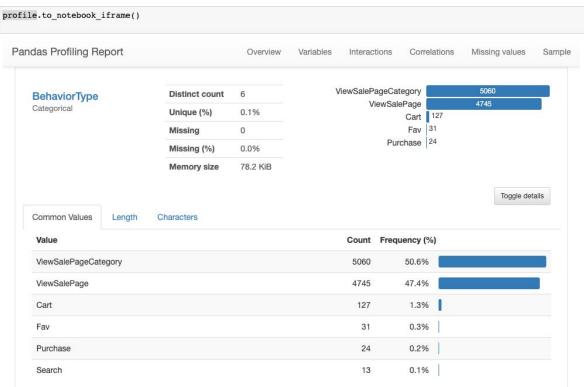
Categorical

Distinct count	6
Unique (%)	0.1%
Missing	0
Missing (%)	0.0%
Memory size	78.2 KiB



Toggle details

行為比例



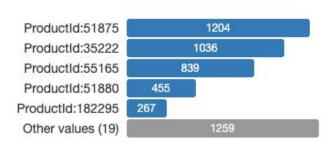
類別比例

Categoryld

Categorical

MISSING

Distinct count	24
Unique (%)	0.5%
Missing	4940
Missing (%)	49.4%
Memory size	78.2 KiB



Toggle details

Hack一下

https://www.so-nice.com.tw/v2/official/SalePageCategory/51875?sortMode=Curator https://www.so-nice.com.tw/v2/official/SalePageCategory/35222?sortMode=Curator

商品價格

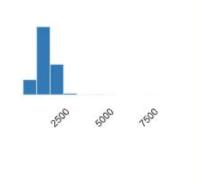
ProductPrice

Real number (R>0)

MISSING

Distinct count	32
Unique (%)	0.7%
Missing	5104
Missing (%)	51.0%
Infinite	0
Infinite (%)	0.0%

Mean	1659.1584967320262
Minimum	390.0
Maximum	7980.0
Zeros	0
Zeros (%)	0.0%
Memory	78.2 KiB



Toggle details

客單價

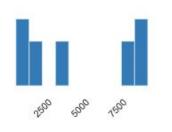
TransactionRevenue

Real number (R>0)

MISSING

Distinct count	12
Unique (%)	50.0%
Missing	9976
Missing (%)	99.8%
Infinite	0
Infinite (%)	0.0%

Mean	4896.25
Minimum	670.0
Maximum	9480.0
Zeros	0
Zeros (%)	0.0%
Memory size	78.2 KiB

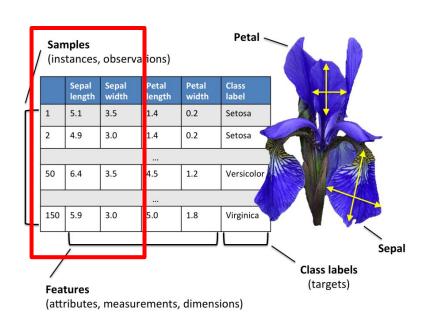


Toggle details

Live Demo 分析

Logistic Regression

資料集- Iris data set



給花瓣、萼片 長跟寬

預測Iris花的品種

資料集- Iris data set

IRIS dataset



Iris Versicolor



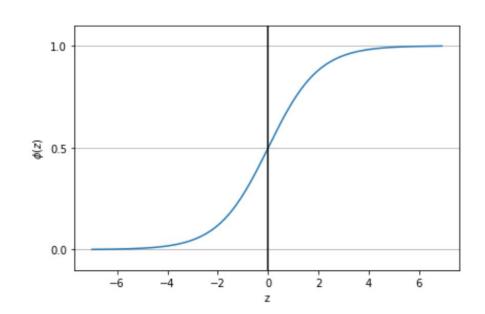
Iris Virginica



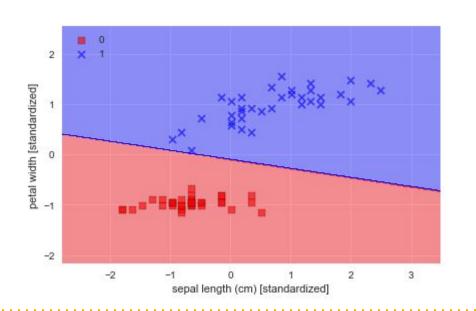
Iris Setosa

Logistic Function

$$z = w^T x$$
$$\phi(z) = \frac{1}{1 + e^{-z}}$$



Logistic Regression 能做什麼? 線性分類



aX+bY+c=0

Live Demo

課後學習資源

學習資源

Blog

[資料分析&機器學習] 第2.3講: Pandas 基本function介紹(Series, DataFrame, Selection, Grouping)

[資料分析&機器學習] 第2.4講: 資料前處理 (Missing data, One-hot encoding, Feature Scaling)

[<u>資料分析&機器學習] 第2.5講: 資料視覺化 (Matplotlib, Seaborn, Plotly)</u>

[資料分析&機器學習] 第3.1講: Python 機器學習以及 Scikit-learn介紹

[<u>資料分析&機器學習</u>] 第3.2講: 線性分類-感知器 (Perceptron) 介紹

[資料分析&機器學習] 第3.3講:線性分類-邏輯斯回歸 (Logistic Regression) 介紹 書

Python 機器學習

學習資源

Github

https://github.com/Avik-Jain/100-Days-Of-ML-Code



Q&A

Thanks

