

Machine Learning Submission Assignment #1

徐玉山
0851915

March 23, 2020

1 Problem 1. Unbiasedness of OLS

Suppose that Population model is:

$$y = \beta_1 + \beta_1 x + u$$

and OLS regression is:

$$\hat{y} = \hat{\beta}_1 + \hat{\beta}_1 x$$

(b) By (0.25)

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x}) u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$E(\hat{\beta}_1) = E(\beta_1) + E\left(\frac{\sum_{i=1}^n (x_i - \bar{x}) u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)$$

$$E(\hat{\beta}_1) = \beta_1 + E\left(\frac{\sum_{i=1}^n x_i * u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right) - E\left(\frac{\sum_{i=1}^n \bar{x} * u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)$$

Since

$$\sum_{i=1}^n x_i * u_i = E(x * u) = 0$$

$$\sum_{i=1}^n \bar{x} * u_i = \bar{x} * n * E(u_i) = 0$$

So

$$E(\hat{\beta}_1) = \beta_1$$

(a) By (b) and (0.9)

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$E(\hat{\beta}_0) = E(\bar{y}) - E(\hat{\beta}_1 \bar{x})$$

$$E(\hat{\beta}_0) = \bar{y} - \bar{x} E(\hat{\beta}_1)$$

$$E(\hat{\beta}_0) = \bar{y} - \bar{x} \beta_1$$

$$E(\hat{\beta}_0) = \beta_0$$

Since

$$\bar{y} = \beta_1 \bar{x} + \beta_0$$

2 Problem 2

(a)

$$\hat{\beta}_1 = -0.3188642859705657$$

$$\hat{\beta}_0 = 32.142711644500494$$

(b) the percentage of students passing the math10 decrease 3.1

(c) 不合理，給午餐錢理論上不是應該考得更好嗎？

(d)

$$\beta_0 = -20.36081647$$

$$\beta_1 = 14.34441037$$

$$\beta_2 = -0.3045853$$

R-square: 0.8200728991486963 (表示其實這份math10 和這兩個參數蠻相關的) 由於expend取了log，導致值不會太大所以我想lnchprg 變化不會太大

3 Problem 3

- (a) 數據(data)僅代表數據本身，例如線上遊戲(英雄聯盟,5v5玩家dota遊戲)中玩家的登入資訊、對戰紀錄。

信息(information)代表分析數據之間的關係得到的東西，可用來回應一些簡單的問題，像是這個遊戲晚上八點的平均在線人數。

知識(knowledge)代表信息的集合吸收之後所整合出的能力，且可以回答更為複雜的問題，例如這個遊戲現在那些角色處於強勢，並分析出該如何平衡角色以及裝備等等。

智慧是一種非確定性的，做正確判斷和決定的能力，包含對知識的最佳使用。智慧可以回答的問題不僅限於前三者所構成的問題。例如從打英雄聯盟的過程中悟出了做人處事的道理之類的。

- (b) supervised learning: 給定輸入和預期輸出，從中學到其中的關係。

例子: 給1000張含有貓並且標記為貓的圖片，1000張含有狗且標記為狗的圖片，希望學到如何辨認貓和狗圖片之間的差異

unsupervised learning:

在非監督式學習裡，會將一組未明確指示處理方式的資料集交給深度學習模型。訓練資料集是一組無特定期望結果或正確答案的例子，神經網路會嘗試擷取出有用特徵並分析其結構，以求自動找出資料結構。

例子: 給1000張不同鳥類的圖片，根據鳥種的羽色、尺寸或嘴喙形狀等線索大致加以區分種類(clustering)

4 Problem 4

- (a) 訓練了一個大型的神經網路，並在ImageNet比賽中取得了最好的結果。

實現利用多GPU來加速訓練

提出了一些方法來提高模型有效性，減少訓練時間(ReLU)

防止過擬合的方法

(1) Dropout: 放棄一些神經元的輸出，使model的輸出不特別依賴某些神經元

(2) 數據增強: 從256*256抽取224*224，並做水平翻轉，一張原始圖片可產生2048倍的訓練圖片。

- (b)

- (1) Defining problem
圖片分類問題，給定一張圖片，輸出它屬於該類別的分數(共一千個類別)
- (2) Gathering data
ImageNet 圖片數量：1500萬
類別數：2.2萬
圖片來源：網上人工標註的
ILSVRC-2010 (2010 ImageNet Large-Scale Visual Recognition) 圖片數量：120萬張訓練圖片，5萬張驗證圖片，15萬張測試
圖片類別數：1000
ILSVRC-2012
與2010相比，它的測試集標籤是不公開的。評價指標：top-1準確率和top-5準確率
下載上述的資料集
- (3) Data preparation
有長寬問題，先取短邊等比例壓縮到256pixels，再取長邊中間的256pixels。
數據增强的部分，從256*256抽取224*224並且做水平翻轉，達到2048倍的訓練量。
- (4) Model development
網路包含八層帶權重層；前五層是卷積層，其餘三層是全連接層。
第一層卷積層過濾維度為224x224x3的輸入影像，
96個大小11x11x3的kernel，stride為4
第二層卷積層並以256個大小5x5x48的kernel做過濾。
第三、四、五卷積層彼此連接，而沒有中介的池化或正規化層。
第三層卷積層擁有384個大小3x3x256的kernels
第四層卷積層有384個大小3x3x192的kernels
第五層卷積層有256個大小3x3x192的kernels
每一層全連接層都擁有4,096個神經元。
最後一層全連接會有1000個輸出，分別對應到1000個類別的分數，並取softmax

- (5) Training
stochastic gradient descent
batch size為128
momentum為0.9
weight decay為0.0005

- (6) Evaluation
ILSVRC2010測試集

Model	Top-1	Top-5
Sparse coding	47.1%	28.2%
SIFT+FVs	45.7%	25.7%
CNN	37.5%	17.0%

- (7) Parameter tuning