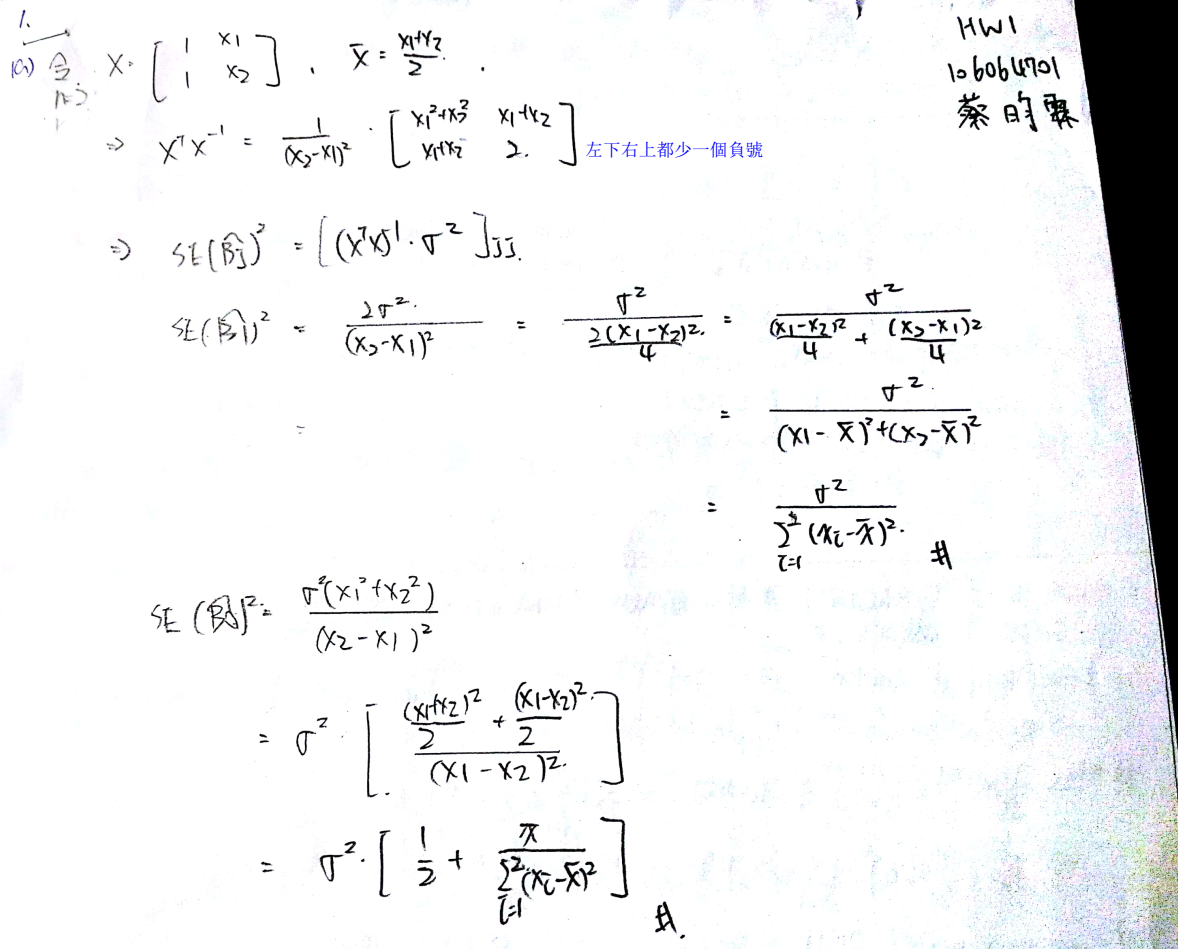
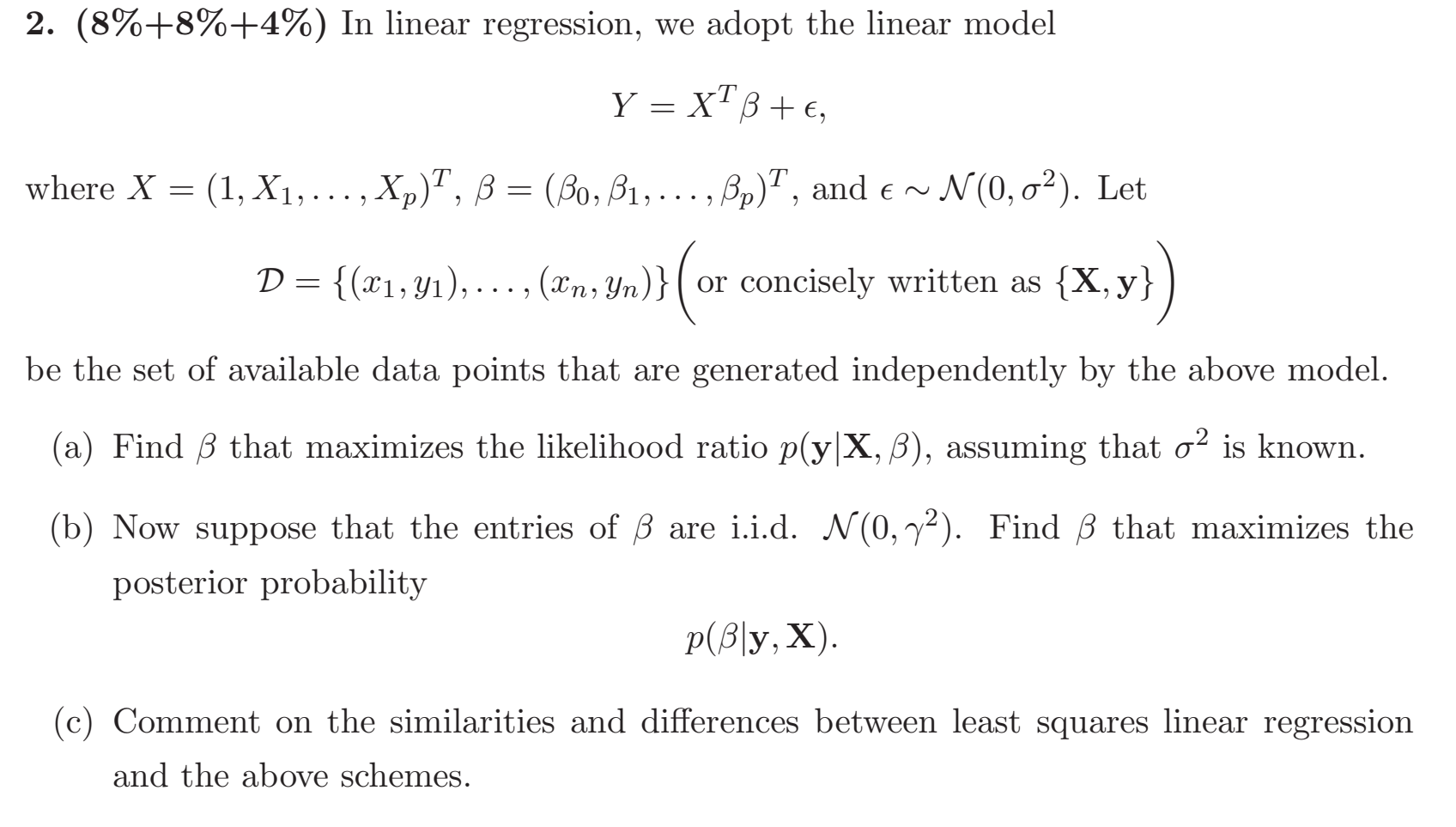
106064701 蔡昀霖

1.



2. 

(a)

, 由題目知 ∼ N (0, σ2)，那麼Y為Normal Distribution 的線性組合，所以Y~(,σ2), 。

🡪所以Likelihood Function:

L()=

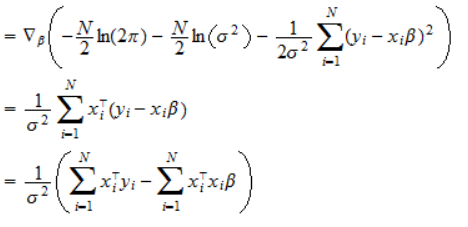
, i:表示觀察的數目

🡪log-Likelihood:

🡪Maximum Likelihood:

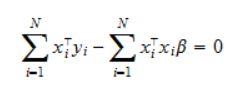
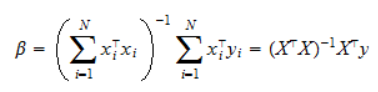
對log-Likelihood最大化等效於對Likelihood做最大化。

求取全微分:



左邊的圖示抓來的，他的xi 是我的xi.transpose!!

🡪“極值”必發生在全微分=0(梯度為零)

 #得解

左邊的圖示抓來的，他的xi 是我的xi.transpose!!

(b)這種算法稱為 最大後驗估計(Maximum-a-Posteriori (MAP) Estimation)

我先將利用貝氏定理化為以下式子

由(a)知式子可以化簡，並且我們要求取B使的後驗機率最大

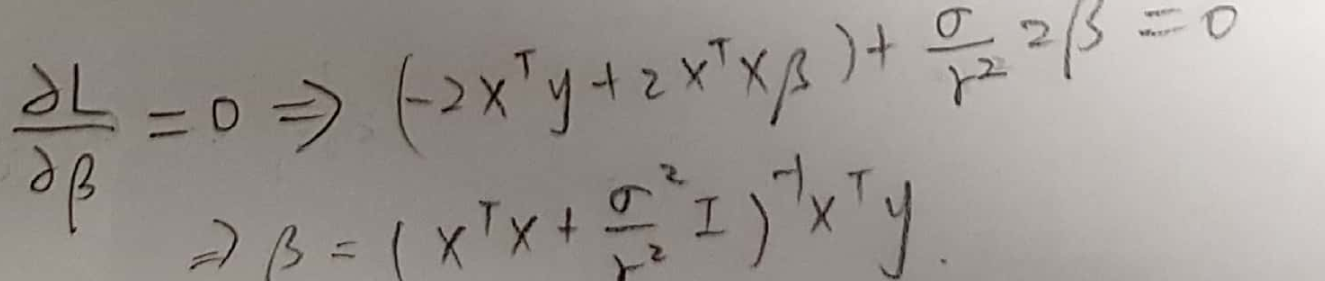
同樣的，我們先對取ln，並由於p(y,X)是一個常數，所以會得到以下等效的式子:

這邊要注意: 意味著在給定Data X下 y在 下發生的機率，這說明了其機率分布為

對做的偏微分得

,則最大值出現在左式為零時。

##

先轉化為矩陣比較好表示

(c)

這邊我分點描述:

1.最小平方估計是利用最小化L2-Norm 求取參數β。

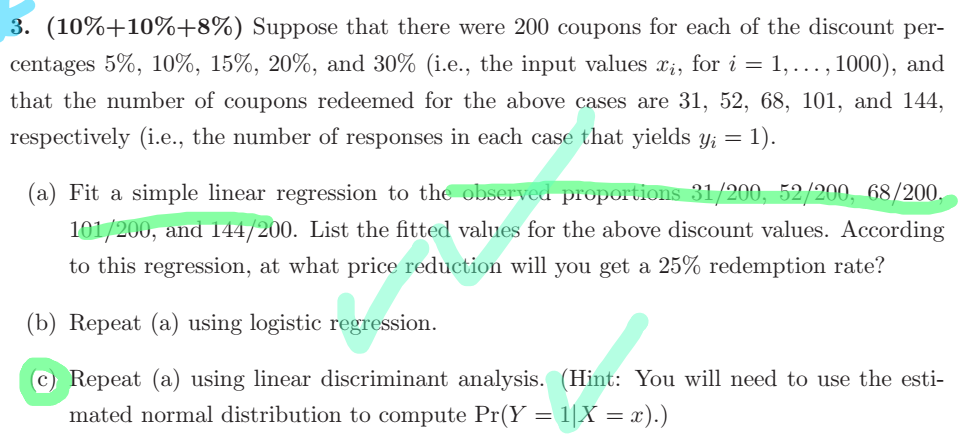
2.最大似然估計則是利用機率的觀點，求取一組參數β使P(D|β)機率最大。

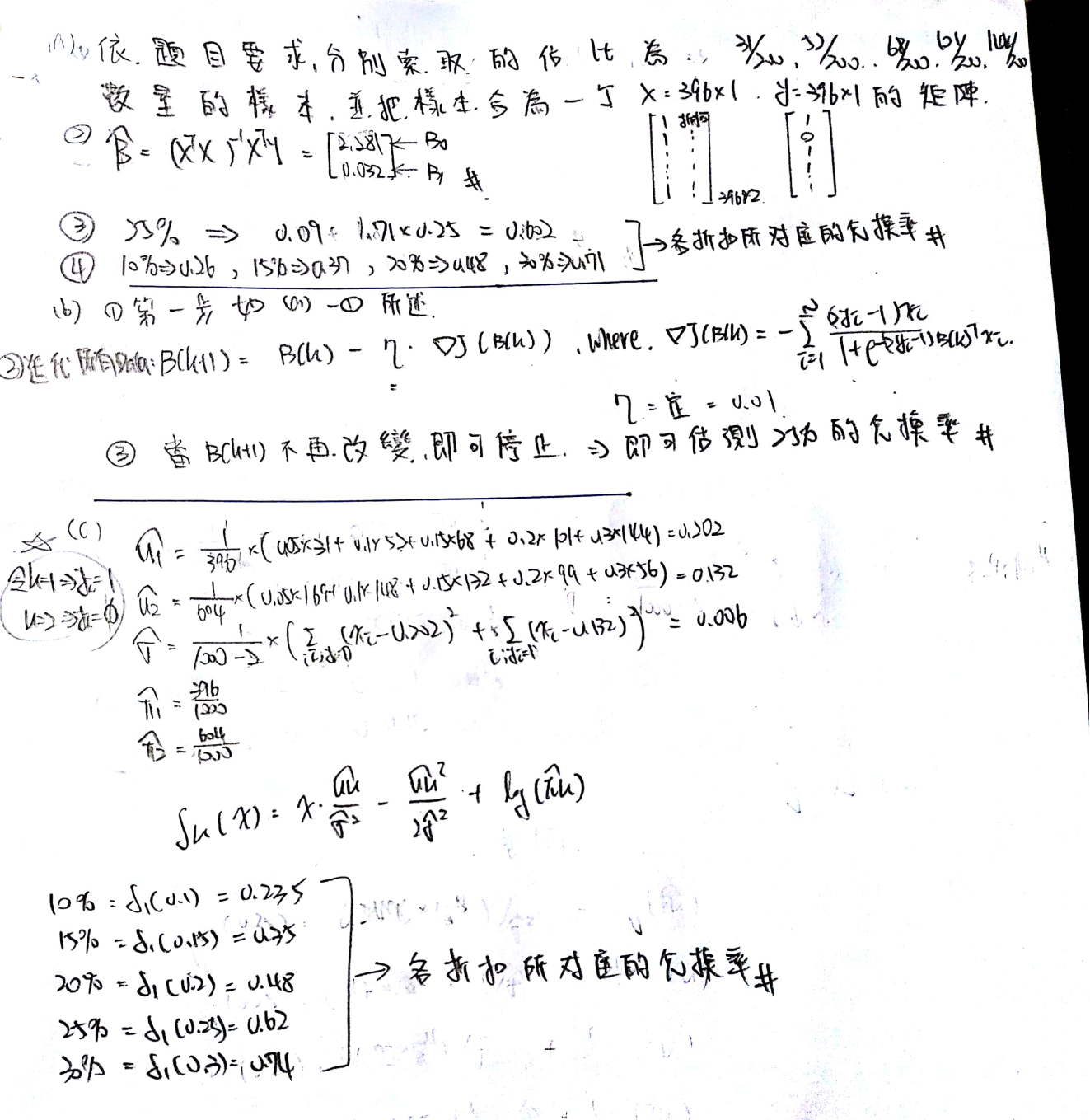
3.當觀測值來自指數族且滿足輕度條件時，最小平方估計和最大似然估計是相同的。

4.最大似然估計(MLE)是求參數β, 使似然函數P(D|β)最大。而最大後驗機率估計(MAP)則是想求β使P(D|β)P(β)最大。求得的β不單單讓似然函數大，β自己出現的先驗概率也盡可能的大。這裡有點神似之後上課應該會上到的正則化概念(懲罰)，只不過一般正則化是使用加法，而這邊使用了乘法，附帶提一點:正則化是避免模型過度擬合(Overfitting)的方法。

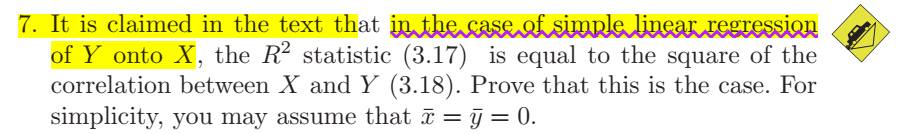
5.最大似然估計(MLE)認為參數本身的機率分布是均勻的(Uniform)，即其機率會是一個常數。

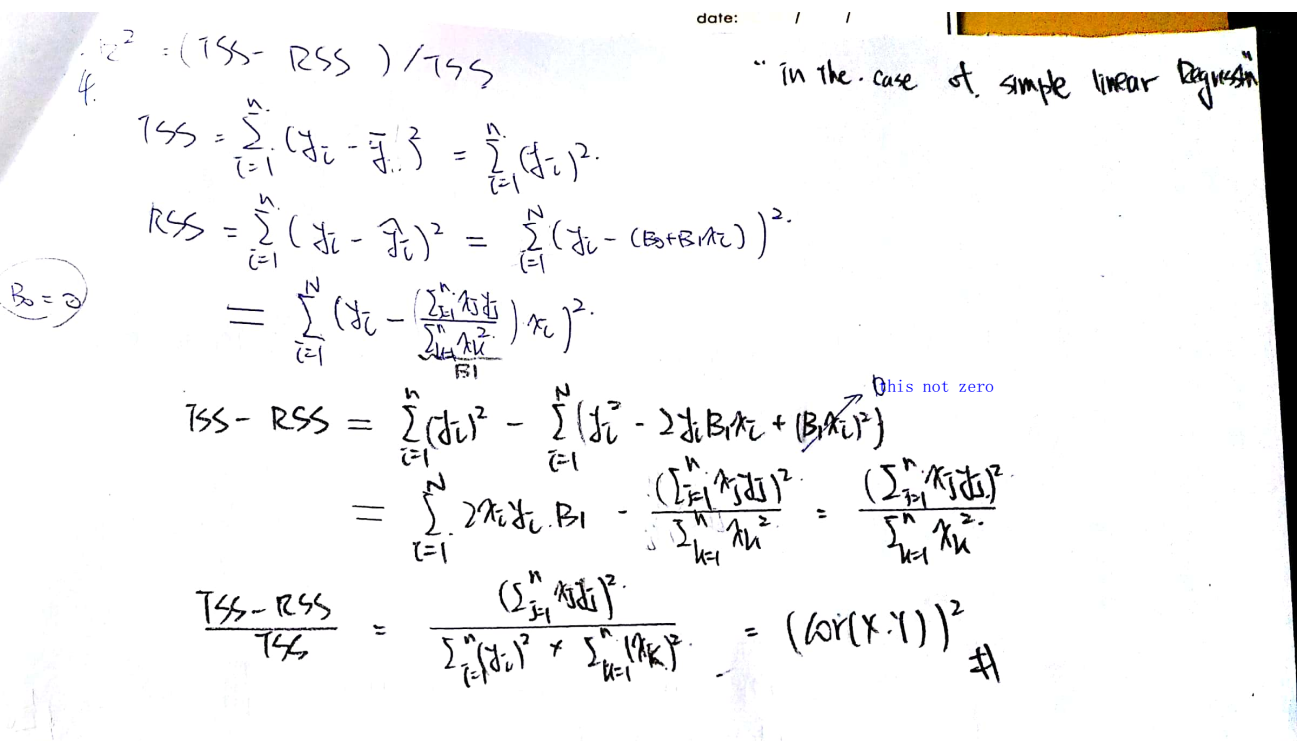


3. 



4.





5.

