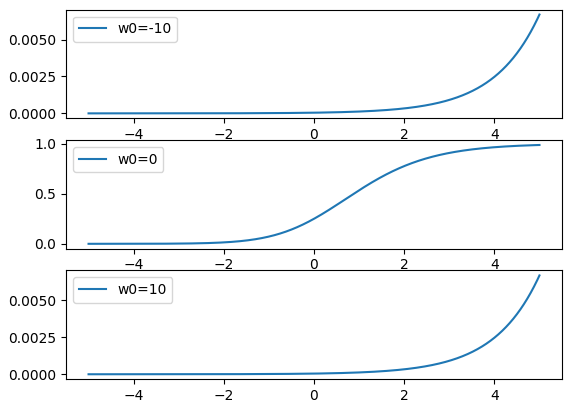
**機器學習 HW2 part2**

**工科系111 E94071209 林政旭**

Question 3.0:

下圖橫軸為w1，設定w0為定值的likelihood圖，由下圖可知不管w0為何，當w1越大時，likelihood function的值皆會越大，並越來越趨近於1，因此，不存在unique的一組w0及w1使得likelihood value有最大值。

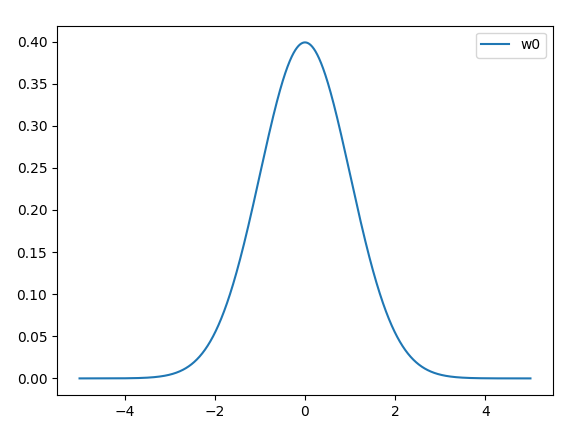


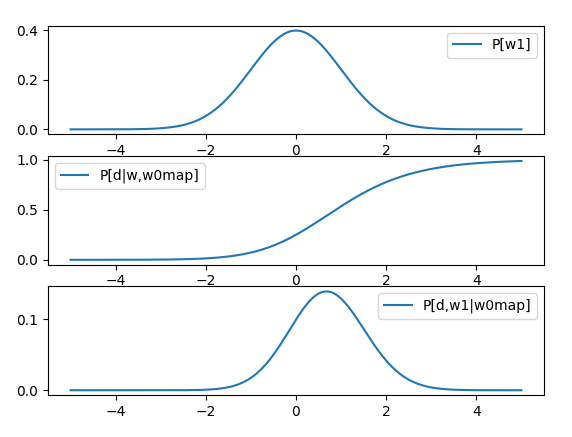
下圖為sigmoid、likelihood、Normal weight prior之P[w] 與Laplace weight prior 之P[w] 的程式圖，並設定x為在-5至5之間的2000個分布點

一張含有 文字 的圖片

自動產生的描述

Question 3.1

1. 下圖橫軸為w0，縱軸為P[w0]之圖，由圖可知最大值發生在w0=0；對其微分可得dP[w0]，另其=0，用程式跑發現最大值發生在-0.0025(非常接近0)，且P[w0]最大值=0.398941



1. 在w0=w0map=-0.002501250625313034時，使用程式跑出來w1最大值發生在0.672836418209104

一張含有 文字 的圖片

自動產生的描述

1. 在w0=w0map=-0.002501250625313034時，posterior mean of w1為

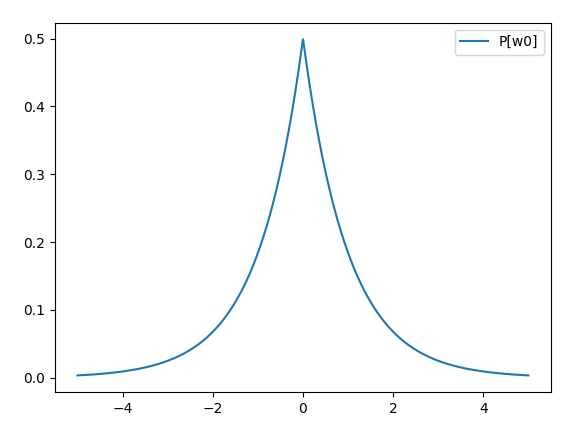
0.7042758569261673

一張含有 文字 的圖片

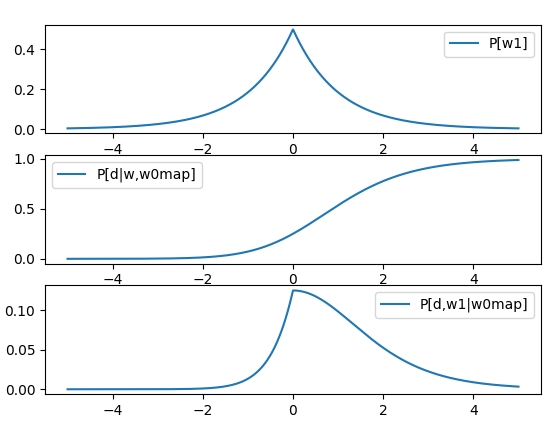
自動產生的描述

Question 3.2:

1. 下圖橫軸為w0，縱軸為P[w0]之圖，由圖可知最大值發生在w0=0；對其微分可得dP[w0]，另其=0，用程式跑發現最大值發生在0.002501250625312146 (非常接近0)，且P[w0]最大值=0.4987509374477932



2.



3. 以分析的角度去求最大值的w1的話，首先可先將P[d,w1|w0map]拆成w1>=0與w1<0兩個部分，又因為w0map約等於0，所以第一部分w1>=0，可將式子約略化簡成，接著再化簡成，因exp(-w1)在0~1之間，所以主要看exp(w1)項，而因為w1越大，exp(w1)越大，整個function值就會越小，所以當w1為0時， function有最大值1/8；接著看第二部分w1<0，可將式子化簡成，接著進一步化簡成，當w1越來越小時，整個function值也會越來越小，而當w1=0時，function有最大值1/8，因此，根據夾擠定理(Squeeze theorem)，P[d,w1|w0map]的最大值為1/8，且發生在0的位置。

下圖為使用程式在w0=w0map=0.002501250625312146時， w1最大值發生在0.002501250625312146，且最大值為0.12499960898510112，約等於1/8

一張含有 文字 的圖片

自動產生的描述

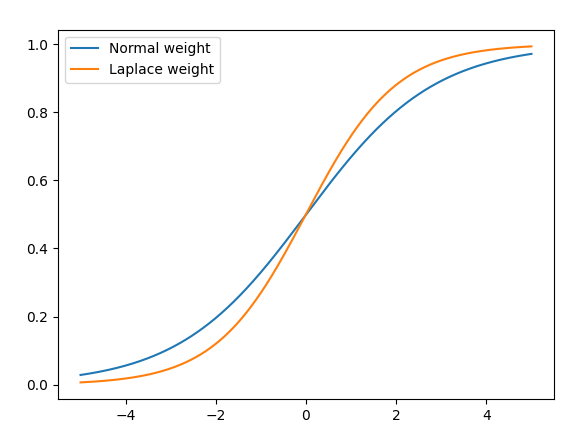
4. 在w0=w0map=0.002501250625312146時，posterior mean of w1為

0.9963922458605604

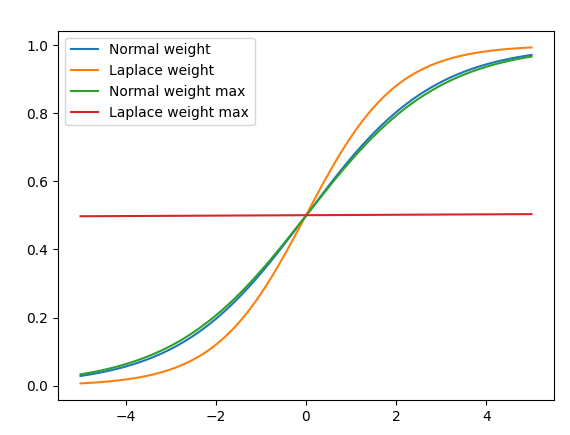
一張含有 文字 的圖片

自動產生的描述

下圖為P[class = Y | 𝑥, 𝑤0, 𝑤1]，其中w0=w0map，w1=posterior mean，由圖中可看到，在x = -1時，Laplace weight prior = 0.270144218660314，Normal weight prior = 0.3303113723034824，皆小於0.5，表示其皆可成功判斷在x = -1時，其有較大的概率不是class Y，而其中，又以Laplace weight prior的表現比Normal weight prior要來的好一些。



下圖的綠色、紅色部分為用w1=w1map來做圖，可以看到Laplace weight prior with w1map 已經變成一條直線了，若使用它來預測class N和class Y將會非常不準確，Normal weight prior with w1map 則是與 posterior mean的w1相差不大



因為P[w1|w0MAP,d] = c\*P[d|w0MAP, w1] P[𝑤1]，所以當P[w1]為不同的prior時，會導致P[w1|w0MAP,d]不同，而導致產生不同的w1map以及不同的posterior mean of w1，預測的精確度也會跟著w1的不同而有所影響

在small datasets的情況下，一般而言標準差是會比較大的，若是所有抽樣的class Y 的datasets的x值皆大於0，那麼我們可能就會有先入為主想法去限定條件如P[class = Y| x <0 ,w]=0，這會使x<0不產生任何機率，對於我們預測其結果來說是不利的，因為我們會損失掉所有x<0的可能性，因此，常態分布很常被使用，因為它可以集中小範圍的數據，且對於任何x值都保有一定的機率，永不歸零。