

ML2016 HW1 Report

R05921040 謝友恆

- Code:

```
g2 = np.zeros(162)
gb2 = 0
g = np.zeros(162)
for i in range(iternum):
    #r is a random number to identify the data
    #chosen in stochastic gradient descent.
    #train is the ndarray of training data,
    #train2 is the stochastic chosen data.
    r = random.randint(0, 5252)
    train2 = train[:, r:r+400]
    #c is the array of (y -sigma(x_i*w_i) - b)
    c = (y[r:r+400] - w.dot(train2) - b)
    #g is the gradient descent error matrix
    #gb is the gradient descent error for b.
    if(reg == 1):
        g = (2 * (c * -1 * train2).mean(axis = 1)) + 2 * theta * w
    else :
        g = (2 * (c * -1 * train2).mean(axis = 1))
    gb = (-2 * c.mean())
    #g2 and gb2 is for adagrad
    g2 = g2 + g*g
    gb2 = gb2 + gb*gb
    #the feedback is as following
    #b is the constant and w is the parameter array
    w = w - alpha * g / np.sqrt((g2+eta).astype(float))
    b = b - alpha * gb / np.sqrt((gb2+eta).astype(float))
```

- Method:

- Feature Extraction: 在本次的作業中使用了兩種版本的 feature，一種是單純使用前九小時的 PM2.5 資料，第二種是使用前九小時的所有 (18 種) 資料。從每個月的第一個小時開始每一小時取一次一直到每個月二十號為止，總共有 5652 筆資料。
- Preprocessing: 而觀察發現資料內會有一些欄位為-1，由於不可能出現濃度-1，因此可能是出現錯誤或沒有值的資料，對每一種類的資料而言，若其九小時內出現-1 的資料，則在這次的實作中將其排除，並補上這九小時內不為-1 的資料的平均。較好的方式應該是對其前後合法的資料做內插，不過並沒有實作這項功能。
- Linear Regression: 本次的 Linear Regression 使用了 stochastic gradient descent，一次取四百筆資料來計算 gradient descent，同時搭配 adagrad 減少 training 時間。

- Regularization:

Regularization 的部分在是在誤差項的計算由

$$L = \sum_i (y - x_i \times w_i - b)$$

改為

$$L = \sum_i (y - x_i \times w_i - b)^2 + \lambda \times \sum_i w_i$$

希望透過這樣的方式減少因雜訊產生的誤差。

適當的 regularization 會有不錯的效果，雖然在 train 的時候 error 會稍大，但更能反映對真實資料的表現，但過大的 regularization 會使 error 變大
實際測試結果如下：

Lambda	0.1	0.5	1	1.5	2
Local error	5.52	5.65	5.71	5.66	5.66
Public set error	5.72	5.91	5.97	5.95	5.87
Private set error	7.19	7.36	7.28	7.42	7.31

- Learning Rate:

在沒有使用 adagrad 的情況下，若讓 learning rate 過大，雖然一開始收斂速度快，但會導致 training 時 error 跳動，無法達到較好的結果，甚至會無法收斂。而使用 adagrad 之後 Learning rate 造成的影響就相對小了，就算用較大的 learning rate 加速前期收斂，後期也會因為分母變大的關係自動變小。