

資料科學導論 Report—HW4

105502035 資工 4B 葉軒

程式說明：

- Loss function
 - $L(\boldsymbol{\theta}) = \frac{1}{2} \sum_{(i,j) \in \tilde{k}} (r_{ij} - \hat{r}_{ij})^2 + \frac{\lambda}{2} \|\boldsymbol{\theta}\|^2$
 - , where $\hat{r}_{ij} = \mu + b_i + c_j + \mathbf{p}_i \cdot \mathbf{q}_j$
- Let $d_{ij} = r_{ij} - \hat{r}_{ij}$, the gradients are
 - $\nabla_{b_i} = -d_{ij} + \lambda b_i$
 - $\nabla_{c_j} = -d_{ij} + \lambda c_j$
 - $\nabla_{\mathbf{p}_i} = -d_{ij} \mathbf{q}_j + \lambda \mathbf{p}_i$
 - $\nabla_{\mathbf{q}_j} = -d_{ij} \mathbf{p}_i + \lambda \mathbf{q}_j$
- Update rule of SGD
 - $\theta^{(k+1)} = \theta^{(k)} - \eta \nabla_{\theta^{(k)}}$

按照老師課堂power point的流程，我將過程換成程式碼，利用param beta(regularization parameter) → lambda來改善overfitting的問題。

結果分析：

Training loss 會隨著 iteration的提高而改善。此外，regularization parameter值過小，loss越多，而當regularization parameter值過大，會發生overflow的問題。

下圖為預設的 learning rate 0.0001(1e-4), regularization parameter 0.01(1e-2)

```
Iteration: 10 ; error = 0.9782
Iteration: 20 ; error = 0.9526
Iteration: 30 ; error = 0.9368
Iteration: 40 ; error = 0.9257
Iteration: 50 ; error = 0.9173 )
```

